

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 12, 2002, 03:22:00 ; Search time 2485 Seconds

(without alignments) 1244.804 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051

Sequence: 1 MLAWALPSILRLGAQETE.....LSPGNQYLHILQNNPHYNSP 191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPRO.spool/US0946265/runat_07112002.101526.24854/app.query.fasta.1.327
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Blousum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0946265 -CGEN 1.1.2024 @runat_07112002.101526.24854 -NCPU=6 -ICPU=3
-NO.XLPXY -NO.MAP -LARGROUPEXT -NEG.SCORES=0 -WAIT -LONGLOS -DEV.TIMEOUT=120
-MAIN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:
2: em_esthum:
3: em_estin:
4: em_estlum:
5: em_estlov:
6: em_estpl:
7: em_estro:
8: em_hlc:
9: gb_est1:
10: gb_est2:
11: gb_hlc:
12: gb_est3:
13: gb_est4:
14: gb_est5:
15: em_estfun:
16: em_estom:
17: gb_gss:
18: em_gss_hum:
19: em_gss_iny:
20: em_gss_pln:
21: em_gss_vrt:
22: em_gss_fun:
23: em_gss_mam:
24: em_gss_mus:
25: em_gss_other:
26: em_gss_pro:
27: em_gss_rod:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	791	75.3	575	10	AM467554
C 2	692	65.8	520	10	AM076051
C 3	692	65.8	589	12	BG174272
C 4	692	65.8	637	13	B1149595
C 5	692	65.8	676	13	B1550838
C 6	692	65.8	682	11	AK008335
C 7	692	65.8	686	14	BQ951956
C 8	692	65.8	688	12	BG244455
C 9	692	65.8	703	13	B1453419
C 10	690	65.7	650	13	B1155774
C 11	687	65.4	610	10	B1154844
C 12	678	64.5	600	12	BG871384
C 13	677	64.4	627	9	AA228200
C 14	675	64.2	645	13	BG975104
C 15	671	63.8	614	10	BE199698
C 16	669	63.7	842	13	B1409815
C 17	668	63.6	1045	12	BE913706
C 18	665	63.3	706	12	BF163190
C 19	658	62.6	619	12	BF302505
C 20	654	62.2	555	12	BF076851
C 21	653.5	60.3	601	9	AA238752
C 22	632	60.1	580	13	B1556032
C 23	620	59.0	575	9	AI507116
C 24	616	58.6	591	9	AA238564
C 25	602	57.3	513	14	BQ566677
C 26	586	55.8	528	9	AA734993
C 27	575.5	54.8	534	9	AA734805
C 28	567	53.9	547	9	AA689693
C 29	553	52.6	548	9	AA597240
C 30	529	50.3	548	9	AI585767
C 31	523	49.8	518	10	BE135520
C 32	521	49.6	438	13	BI849729
C 33	510	48.5	384	10	BE654595
C 34	503	47.9	409	12	BG187105
C 35	489	46.5	564	9	AA499942
C 36	481	45.8	510	12	BF152060
C 37	477	45.4	665	13	BG969181
C 38	463	44.1	384	9	AA689633
C 39	462	44.0	573	9	AA543570
C 40	411	39.1	451	12	BF151994
C 41	405.5	38.6	649	9	AI108979
C 42	395.5	37.6	625	10	AM941788
C 43	394.5	37.5	604	13	B1504243
C 44	385.5	36.7	649	13	B1504245
C 45	382	36.3	472	13	B1291182

ALIGNMENTS

RESULT 1
LOCUS AM467554/C
DEFINITION he21d02.x1 NCI CGAP CML1 Homo sapiens cDNA clone IMAGE:2919651 3'
similar to TR:O75594 O75599 PEPTIDOGLYCAN RECOGNITION PROTEIN
PRECUSOR. // mRNA sequence.
ACCESSION AM467554.1 GI:7037660
VERSION AM467554
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS NCI-Cgap <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bdip/image/image.html
Seq primer: -40UP from GIBCO
High quality sequence stop: 359.
Location/Qualifiers
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2919651"
/clone_lib="NCI CGAP CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: PCMV-SPORT6; Site: 1;
Salt: Site 2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Library constructed by Life Technologies."
DB: 10 147 c 178 g 135 t 5 others

BASE COUNT 110 a 147 c 178 g 135 t 5 others
ORIGIN

Alignment Scores:

Pred. No.: 6.75e-77 Length: 575
Score: 791.00 Matches: 147
Percent Similarity: 93.67% Conservative: 1
Best Local Similarity: 93.04% Mismatches: 10
Query Match: 75.26% Indels: 1
DB: 10 Gaps: 0

US-09-462-625-4 (1-191) x AM467554 (1-575)

QY 34 TTPYSAALAEUASERGLUCYSAAGLHLSERLEUPROLEUAAGTYRVAVAL 53
|||||
DB 575 TGAAGGCCCGGCGATCAGATGCGCCGACGACTGGAGCCTTACCTATGTGGTG 516
QY 54 ValSerHisThrAlaGlySerSerCysAsnThrProAlaSerCysGlnGlnAlaArg 73
||| |||
DB 515 GATATGCACCGGCGGAGCAGACAGTGAACACCCCGCTTG -TGCAGCAAGCAGCCCGG 457
QY 74 AsnValGlnHisThrMetLeuThrLeuGlyTTPCysAspValGlyTyrAsnPhelau 93
|||||
DB 456 ATGTCCAGACTACCACTAGAAAGACATGGGCTGGTGCACCTGGCTCAACTTCCTG 397
QY 94 IleGlyLysAspLeuValTyrGlnGlyArgGlyTyrAsnPherHrgValAlaHisSer 113
|||||
DB 396 ATTGAGAGAGAGCGGCTCGATACGAGGCGCGCTGGAACTTCAGGGTGGCCCTCA 337
QY 114 GlyHisLeuTyrAsnProMetSerIleGlyIleSerPheMetGlyAsnTyrMetAspArg 133
336 GGGCACTTATGGAACCCATGTCATTCATTCAGCTCATGGCAANTGATGATCGG 277
QY 134 ValProThrProGlnAlaIleArgAlaIleGlnGlyLeuLeuAlaGlyValAlaGln 153
|||||
DB 276 GTGCCACACACCCAGGCGATCCGGGAGCCCGAGGGTTATGGCTGCGGGCTCGAG 217
QY 154 GlyAlaLeuArgSerAsnTyrValLeuLysGlyHisArgAspValGlnArgThrLeuSer 173
DB 216 GGAGGCCCTAAGTCCAACTATGTGTCAAAGACACCGGAGTGTGAGGTACACTCTCT 157
QY 174 ProGlyAsnGlnLeuTyrHisLeuIleGlnAsnTyrProHisTyrArgSerPro 191
DB 156 CCAGGCAACACCTTATACCTCATCCAGAAATGGCCACANTACCGCTCCGCC 103

RESULT 2
AM076051/c 520 bp mRNA linear EST 13-OCT-1999
LOCUS
DEFINITION
x883901.x1 NCI_CGAP_CML1 Homo sapiens CDNA clone IMAGE:2573424 3'
similar to TR:O75594 O75594 PEPTIDOLYCAN RECOGNITION PROTEIN
PRECUSOR, mRNA sequence.
AM076051
ACCESSION
AM076051.1 GI:6031049
VERSION
AM076051.1
KEYWORDS
EST.

ORGANISM

human.

REFERENCE

1 (bases 1 to 520)

AUTHORS

NCI-CGAP

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bdip/image/image.html
Seq primer: -40UP from GIBCO
High quality sequence stop: 252.
Location/Qualifiers
1..520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2573424"
/clone_lib="NCI CGAP CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: PCMV-SPORT6; Site: 1;
Salt: Site 2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Library constructed by Life Technologies."
DB: 10 139 c 160 g 108 t 6 others

FEATURES

source

BASE COUNT 107 a 139 c 160 g 108 t 6 others
ORIGIN

Alignment Scores:

Pred. No.: 4.58e-66 Length: 520
Score: 692.00 Matches: 129
Percent Similarity: 89.04% Conservative: 1
Best Local Similarity: 88.36% Mismatches: 15
Query Match: 65.84% Indels: 1
DB: 10 Gaps: 0

US-09-462-625-4 (1-191) x AM076051 (1-520)

QY 47 LeuProLeuArgTyrValValValSerHisThrAlaGlySerSerCysAsnThrProAla 66
|||||
DB 520 CTGGCGCTTACCGTATGTGGTGTGATGGAAGACAGGCTGGCCAGTACGCTACACCTCGCCT 461
QY 67 SerCysGlnGlnAlaIleArgAsnValGlnHisTyrHisMetLeuThrLeuGlyTTPCys 86
|||||
DB 460 TTGTGCACACACAGGCGCGAATGTGTACACTTCACATGAAACACTGGGTTGGTGC 401
QY 87 AspValGlyTyrAsnPhelauIleGlyLysAspGlyLeuValTyrGlnGlyArgGlyTyr 106
DB 400 GACGTGGCTTAACACTTCCTGATTTGAGAAAGACGGGCTCGTATACGAGGCGCTGGCTG 341
QY 107 AsnPherHrgValAlaHisSerGlyHisLeuTyrAsnProMetSerIleGlyIleSerPhe 126
DB 340 AACTTCAGGCGGCGCCACTCAGAGTCACTTATGGAACCCCATTCATTCGCACTACGCTTC 281
QY 127 MetGlyAsnTyrMetAspArgValProThrProGlnAlaIleArgAlaIleGlnGlyLeu 146

Email: cgaab5-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Incyte Genomics, Inc.
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0254 row: n column: 17
High quality sequence stop: 570.

Location/Qualifiers
1. .589

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="4457680"
/clone_id="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DHI08"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Saliv
Site_2: Notti; Cloned unidirectionally. Primer: Oligo dT
Library constructed by Life Technologies. Investigator
Providing samples: Gilbert Smith, NIH"
174 c 161 g 128 t

```

5.57e-66	Length:	589
692.00	Matches:	127
78.318	Conservative:	21
67.208	Mismatches:	33
65.848	Indels:	8
12	Gaps:	1

1) x BG174272 (1-589)

aTrrpalaLeuProSerLeuLeuArgLeuGlyalaGlnIuThrglu 20
 1 ||||| ::||| ||
 CTGTCCTCCCTGGCCTCCCTGGCTCGCAACCTCC----- 54

QY 21 AspProAlaCysGlySerProIleValProAlaGlnGluTrpLysAlaLeuAlaSerGlu 40
 Db 55 -----TGCAGTTTCATGTCGCCCCGGACAGTGAAGGAGGGCCCTGGCATCCGAG 102
 QY 41 CysAlaGlnHisLeuSerLeuProLeuAlaGlyValValValSerHisThrAlaGlySer 60
 Db 103 TCTCTAGCCGCGCTGGGGACCCAGTTCGCTACGCTGTATCTCACACACAGCCGGCAGC 152
 QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHisMet 80
 Db 163 TTTCTGCACACAGCCCGGACCTCTGTGTAAACACAGGCCCCGGAATGTGAGCATTTCCACAG 222
 QY 81 LysThrLeuGlyTTPCysAspValGlyTyrAsnPheLeuIleGlyLysLysPGLysLeuVal 100
 Db 223 AATGAGCTGGGCTGGTGGCATGTAGCTCAACTCTTATTGGAAAGGACGGTCAATGTC 252
 QY 101 TyrGlnGlyArgGlyTyrAsnPheThrIleArgLysHisSerGlyHisLeuTrpAsnProMet 120
 Db 283 TATGAAGCCGAGGCTGGACATCAAGGGTGCACACACAGGGCCCCATCTGGAAATCCCATG 342
 QY 121 SerIleGlyLysSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 Db 343 TCTATTGGCATACCTTCATGGGGAACCTCATGAGACCGGGTACCCGCAACGGGGCCCTC 402
 QY 141 ArgAlaIleGlnGlyLeuLeuAlaLysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
 Db 403 CGTGTGCCCCCTAAATCTTCTGTGAATGATGGGGTGTCTCGGGGCTTCTTGGATCCAACTAT 462
 QY 161 ValLeuLysGlyHisArgAspValGlnIleArgThrLeuSerProGlyIleAsnGlnLeuThrHis 180
 Db 463 GAAGTCAAAGGACACCGGGATGTGCAAAAGCACTCTCTCCACGGTGCACCAACTATATCAG 522
 QY 181 LeuIleGlnAsnTrpProHisTyrArg 189
 Db 523 GTCACTCAAAAGCTGGAAACTATCCGA 549

LOCUS	B114595	637 bp	mRNA	linear	EST 05-JUL-2001
DEFINITION	602848402F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5011897 5',				
ACCESSION	mRNA sequence.				
VERSION	B114595				
KEYWORDS	B114595.1	GI:14609596			
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 637)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D.				

Close distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: LLAM11062 row: C column: 02
 High quality sequence step: 632.
 Location/Qualifiers
 1..637

```

/clone="IMAGE:5011897"
/clone_lib="NCI CGAP Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"

```



```

DB      478 GAAGTCAAGACACCGGATGTCAGAACACTCTCTCCAGGTACCAACTCTATCAG 537
OY      181 Leu1leu1aasatp1proh1st1y1atg 189
DB      538 GTCATCCAAAGCTGGACACCTACCGA 564

RESULT 6
AK008335
LOCUS   AK008335
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length
           enriched library, clone:2010107E11:peptidoglycan recognition
           protein, full insert sequence.
ACCESSION AK008335
VERSION   AK008335.1 GI:12842459
KEYWORDS  HTC; CAP trapper.
SOURCE    Mus musculus (strain: C57BL/6J) adult male small intestine cDNA to
           mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library
           clone:2010107E11.
ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS  Carninci, P. and Hayashizaki, Y.
TITLE    High efficiency full-length cDNA cloning
JOURNAL  Meth. Enzymol. 303, 19-44 (1999)
MEDLINE  99279253
PUBMED   10349636

REFERENCE
AUTHORS  2
TITLE    Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
JOURNAL  Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
MEDLINE  Prepare full-length cDNA libraries for rapid discovery of new genes
PUBMED   Genome Res. 10 (10), 1617-1630 (2000)
         11042159

REFERENCE
AUTHORS  3
TITLE    Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
JOURNAL  Konno, H., Akiyama, J., Nishii, K., Kitsuana, T., Tashiro, H., Itoh, M.,
MEDLINE  Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
PUBMED   Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
         Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Obara, E., Watahiki, K.,
         Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,
         RIKEN Integrated sequence analysis (RISA) system--384-format
         sequencing pipeline with 384 multicapillary sequencer
         genome Res. 10 (11), 1757-1771 (2000)
         20530913
         11076861

REFERENCE
AUTHORS  4
TITLE    Kawai, J., Shitagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
JOURNAL  Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
MEDLINE  Aizawa, K., Izawa, M., Nishii, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
PUBMED   Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
         Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
         Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
         Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
         Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M.,
         Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
         Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
         Carninci, P., de Bonaudo, M. F., Brownstein, M. J., Bulc, C.,
         Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
         Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
         Marchionni, L., Mashima, U., Mazzarelli, J., Mombarts, P., Nordone, P.,
         Ring, B., Ringwald, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
         Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
         Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
         Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,
         and Hayashizaki, Y.
         Functional annotation of a full-length mouse cDNA collection
         Nature 409 (6821), 685-690 (2001)
         21085660
         11217851

```

```

REFERENCE
AUTHORS  5 (bases 1 to 682)
          Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
          Aikawa, T., Balderelli, R., Bono, H., Brownstein, M., Bulc, C.,
          Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hasegawa, T.,
          Hara, A., Hayatsu, N., Hill, D., Hiroto, K., Hirooka, T., Hori, F.,
          Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
          Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koyu, S.,
          Kurihara, C., Matsuyama, T., Miyazaki, Y., Nishii, K., Nomura, K.,
          Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
          Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
          Schriml, L., Shibata, K., Shibata, Y., Shitagawa, A., Shiraki, T.,
          Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
          Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yamataka, I.,
          Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
          Hayashizaki, Y.
TITLE    Direct Submission
JOURNAL  Submitted (10-JUL-2000) Yoshitake Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
          URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222,
          Fax: 81-45-503-9216)
          Please visit our web site (http://genome-gsc.riken.go.jp/) for
          further details.
FEATURES
SOURCE    Location/Qualifiers
           1..682
           /organism="Mus musculus"
           /strain="C57BL/6J"
           /db_xref="PANTOM-DB:2010107E11"
           /db_xref="MGD:MGI:1910652"
           /db_xref="taxon:10090"
           /clone="2010107E11"
           /sex="male"
           /tissue_type="small intestine"
           /clone.lib="RIKEN full-length enriched mouse cDNA library"
           /dex_stage="adult"
           1..682
           /gene="Pg1y1p"
           38..586
           /gene="Pg1y1p"
           /note="data source:MGD, source key:MGI:1345092,
           peptidoglycan recognition protein
           putative"
           /codon_start=1
           /protein_id="BAB25611.1"
           /db_xref="GI:12842460"
           /db_xref="MGD:MGI:1345092"
           /translation="MLPACALMLDLGLATSCSFTVPSREMRALPSSCSSRLGHPVRY
           VISHHGSFNSPDSCEQQRANVOHHRKELGNCDAVNFLLIGDGVYVGRGWNIG
           DHRGPIWNPISIGITFMGNFMDRVPKRALRALNLNLECGVSRSGLRSNVEVGHGRDY
           QSLSPGDDLYQVIOQSEHYRE"
           666..671
           /gene="Pg1y1p"
           /note="putative"
           682
           /gene="Pg1y1p"
           /note="putative"
BASE COUNT 147 a 217 c 175 g 143 t
ORIGIN
polyA_site
polyA_signal
Alignment Scores:
Pred. No.: 7 02e-66 Length: 682
Score: 692.00 Matches: 127
Percent Similarity: 78.31% Conservative: 21
Best Local Similarity: 67.20% Mismatches: 33
Query Match: 65.84% Indels: 8
DB: 11 Gaps: 1
US-09-462-625-4 (1-191) x AK008335 (1-682)
OY      1 Metleu1eua1aTTPa1aLeuProSer1euleuArg1euc1y1a1a1a1a1a1a1uThrc1u 20

```

Db		38	ATGTTGTTCCTGGTGCATCCTCCCTCCCGTGGGCTTGCAACCTCC-----	85
Oy		21	ASPTROALAcysSerProIleValProArgansglutPrlyAlaLeuAlaSerGlu	40
Db		86	-----TGCAGTTTCATCTCCGCCCGCACATGATGAGGCGCCCTGCCATCCAG	133
Oy		41	CysalaglnHisLeuSerLeuProMetLgTYrValValSerHisThralaIyser	60
Db		134	TGCTTACCAGCCCTGGGACCCACTTGGCTACGTGTGATGTTCCACACACAGCGGACG	193
Oy		61	SerCysasnThrProAlaSerCysglnglnglnAlaArgasVnaIglHisTyrlsmet	80
Db		194	TTCCTACACAGCCCGGACCTCTGTAAACAGCAGCGCCCGCAATGTGCAGATTACACAAG	253
Oy		81	LysThrLeuglYTrPCysAspValGlYyrAsnPhelEuilelglyGlnaspGlyLeuVal	100
Db		254	AATGACGCTGGGTGGTGGATGTACCTTACCAACTCTTATTGGAGAGACGCTCATGTTC	313
Oy		101	TYrGlnGLYArGGLYTPRAsnPhenrglYlaHIsSserGlnHisLeuTPrasProwet	120
Db		314	TATGAAGCGCGAGGTGGAGAACATCAAAGGTGACCAACAGGGCCCATCTGGATCCCATG	373
Oy		121	SerIleGlyIleSerPheMetGlysantYmeLaSPARgvalProThrProGlnAlaIle	140
Db		374	TCTATTGGCATTCACCTTCATGGGAACTTCACAGAACCGGGTACCCGCAAGGGGCCCCC	433
Oy		141	ArgAlaIagInglyLeuLeuAlaCysglYvaIalaIngInlyAlaLeuArSerAnsTyrr	160
Db		434	CGTGTGCCCTTAACCTTCTGTGAATGTGGAGTGGGGGTCTCTGGGGGCTCTCGATCCAATAT	493
Oy		161	ValLeuLySGLYHISARGspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrlHS	180
Db		494	GAAAGCAAGAAGCACCGGGATGTGCAAGACACTCTCTCCAGGTGAGACCAACTATCAG	553
Oy		181	LeuIleGlnAsnTrpProHistTyrrg	189
Db		554	GTCAATCAAAAGCTGGGAACACTTACCGA	580
RESULT 7				
LOCUS	B0951956		686 bp	mRNA linear EST 21-AUG-2002
DEFINITION	AGENSCOURT 8951306 NCI_CGAP.Co24 Mus musculus cDNA clone			
IMAGE:	6474853 5'			mRNA sequence.
ACCESSION	B0951956			
VERSION	B0951956.1			GI:22367434
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	1 (bases 1 to 686)			
JOURNAL	NIH-MGC http://mgc.ncl.nih.gov/.			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1995)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-remail.nih.gov			
	Tissue Procurement: The Cepko Laboratory			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LMN at:			
	http://image.llnl.gov			
	Plate: LAM14012 row: k column: 14			
	High quality sequence stop: 589.			
FEATURES				
source	Location/Qualifiers			
	1..686			
	/organism="Mus musculus"			
	/strain="FVB/N"			
	/db_xref="taxon:10090"			
	/clone IMAGE:6474853"			
	/clone_11b="NCI_CGAP.Co24"			

[illegible]

OY	61	SerCySanthrProAlSerCysGlnGlnGlnAlaArgAsnValGlnHisTyrHisMet	80
Dd	201	TTCGTGCACAGCCCGGACTCTCTGTGAACAGCAGGCCCGCATGTGCATTTACCACAAG	260
OY	81	IystrLauGLyTYPCysAspValGlyTyraSnpheLeuIleGlyLaSpLyLeuVal	100
Dd	261	AATGAGCTGGGTGGTGCGATGTACCTCAACACTTCCTATTGAGAGAGCGGTGATGTC	320
OY	101	TyTGlUGLyAryGlyTYTPAsnPheThrGlyAlaHisSerGlyHisLeutPAsnPromet	120
Dd	321	TATGAAAGCCGCGAGGTGGAAACATCAAGGGGTGCACACAGGGCCCATCTGGAAATCCATG	380
OY	121	SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle	140
Dd	381	TCTATTGGCATCACTTCTCATGGGAACTTCATGACCGGGTATCCCGCAAAGCGGCCCTTC	440
OY	141	ArgAlaalaGInglyLeuLeuValAcCysgLyAlaAlaagInglyAlaLeuArgSerAnTyr	160
Dd	441	CGTGTCGCCCTTAANCTTCTGTGAAGTGGGGGTCTCTCGGGGCTTCTTGATGCACACTAT	500
OY	161	ValLeuLySgLyHisArgAspValGlnArghrLeuSerProGlyAsnGlnLeutyrHis	180
Dd	501	GAAATCAAAAGGACACCGGGATGTGCAAGAAGCACTCTCTCCAGGTGACCACACTATCATG	560
OY	181	LeuIleGlnAsnTrpProHisTyrArg	189
Dd	561	GTCATCCAAAGCTGGGAACACTACACGA	587
RESULT 10	BII55774	650 bp	mRNA linear EST 05-JUL-2001
LOCUS	BII55774		
DEFINITION	602903433[1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5033867 5'		
ACCESSION	BII55774		
VERSION	BII55774.1		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 650)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Stransberg, Ph.D. Email: cs9apbs-r@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA library arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.lnl.nih.gov Plate: LHM11094 row: f column: 12 High quality sequence stop: 637.		
FEATURES	Location/Qualifiers		
source	1..650		
	/organism="Mus musculus"		
	/strain="129,C57BL/6J,FVB/N"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:5033867"		
	/clone_11b="NCI_CGAP_Mam3"		
	/tissue_type="tumor, gross tissue"		
	/dev_stage="10 months"		
	/lab_host="DH10B"		
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."		
BASE COUNT	137 a 199 c 179 g 135 t		
ORIGIN			

[illegible]

QY 88 ValGlyTyrAsnPheLeuIleGlyGluAspGlyLeuValTyrGluArgGlyTyrPasn 107.
DB 404 GTAGCCTACACTTCCTTATGTGAGAGACGGTCATGCTATGTAAGCGGAGGCTGGAAC 345
QY 108 PheThrGlyAlaHisSerGlyHisLeuTyrPasnPrometSerIleGlyIleSerPheMet 127
DB 344 ATCAAGGGTGACCCACACAGGGCCCATCTGGAATCCCATGCTATGTGGCATCACCTTCATG 285
QY 128 GlysAsnTyrMetAspArgValProThrProGlnAlaIleArgAlaAlaGlnGlyLeuLeu 147
DB 284 GGGAACCTTCATGTGACCGGGGTACCCGCAAAAGCGGGCCCTCCGCTGCGCTAAATCTTCTG 225
QY 148 AlaCySGlyValAlaGlnGlyAlaLeuArgSerAsnTyrValLeuLysGlyHisArgAsp 167
DB 224 GAATGTGGGGGTGTCTCGGGCTTCTTGAGATCCAACTATGCAAGTCAAGACACCGGAT 165
QY 168 ValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHisLeuIleGlnAsnTyrProHis 187
DB 164 GTGCAAAAGCACTCTCTCCAGGTGACCAACTCTATCAGGTCAATCCAAAGCTGGGAACAC 105
QY 188 TyrArg 189
DB 104 TACCGA 99

Search completed: November 12, 2002, 05:42:41
Job time : 2496 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 00:42:09 : Search time 2472 Seconds
(without alignments)
4704.029 Million cell updates/sec

Title: US-09-462-625-3

Perfect score: 718
Sequence: 1 ctgagttacggcgccagag.....ctcctctccataaagatg 718

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estlum: *
5: em_estlov: *
6: em_estpl: *
7: em_estro: *
8: em_hlc: *
9: gb_estl: *
10: gb_estl2: *
11: gb_hlc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_ylt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_trod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	507.8	70.7	575	10	AM467554 he21d02.x
2	467.2	65.1	520	10	AM076051 AM076051 xa83g01.x
3	350.8	48.9	409	11	BG187105 RST6088 A
4	316.6	44.1	682	12	AK008335 Mus muscu
5	316	44.0	703	13	B1453419 B03170810
6	315	43.9	589	12	BG174272 602334571

7	315	43.9	610	13	B1154844
8	315	43.9	637	13	B1149595
9	315	43.9	650	13	B1155774
10	315	43.9	676	13	B1650838
11	315	43.9	686	14	B0951936
12	314.6	43.8	688	12	BG244455
13	314.6	43.8	842	13	B1409815
14	314	43.7	627	9	AA228300
15	306.6	42.7	645	13	BG975104
16	306	42.6	614	10	BE199698
17	305.4	42.5	600	12	BG871384
18	304.8	42.5	555	12	BF076851
19	303	42.2	619	12	BF302805
20	303	40.8	706	12	BF163190
21	293	40.8	1045	12	BE913706
22	292.8	40.8	580	13	B1556032
23	292.2	40.7	575	9	A1507116
24	285.8	39.8	601	9	AA238752
25	277	38.6	513	14	BQ564677
26	274.2	38.2	591	9	AA238564
27	261.8	36.5	528	9	AA734893
28	259.4	36.1	547	9	AA689693
29	256.6	35.7	534	9	AA734805
30	252.8	35.2	548	9	A1585767
31	247.8	34.5	548	9	AA597240
32	240.6	33.5	384	10	BE654595
33	239	33.3	438	13	BE849729
34	236.4	32.9	518	10	BE135520
35	221	30.8	510	12	BE152060
36	215.8	30.1	384	9	AA689633
37	202.6	28.2	573	9	AA543570
38	201.4	28.1	564	9	AA499842
39	196.4	27.4	238	12	BE901379
40	180.8	25.2	451	12	BE151994
41	176	24.5	504	12	BE548605
42	175.4	24.4	472	13	B1291182
43	175.4	24.4	506	9	AA875213
44	169.2	23.6	665	13	BG969181
45	155.4	21.6	384	9	AA015317

ALIGNMENTS

RESULT 1
AM467554/C 575 bp mRNA linear EST 24-FEB-2000
LOCUS he21d02.x1 NCI CGAP CML Homo sapiens cDNA clone IMAGE:2919651 3'
DEFINITION similar to TR:O75594 O75594 PEPTIDGLYCAN RECOGNITION PROTEIN
PRECUSOR. ; mRNA sequence.

AM467554 GI:7037660

ACCESSION

AM467554

VERSION

AM467554.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 575)

AUTHORS

NCI-CGAP

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,

Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/dbp/image/image.html

Seq primer: -40UP from Gibco

Db	40	AAAAACCCACTGCTCTCTTCACATAAAGATG	7
RESULT 3			
LOCUS	Bg187105	409 bp	mRNA linear EST 21-APR-2001
DEFINITION	Bg187105	Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.	
ACCESSION	Bg187105		
VERSION	Bg187105.1	GI:13708792	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 409)		
AUTHORS	Harrington, J. J., Sherif, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J. and Ducar, M.		
TITLE	Creation of genome-wide protein expression libraries using random activation of gene expression		
JOURNAL	Nat. Biotechnol. 19 (5), 440-445 (2001)		
MEDLINE	2122751		
COMMENT	Contact: Scott J. Cain. Athersys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scaine@atersys.com High quality sequence stop: 409. Location/Qualifiers 1. 409 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Athersys RAGE Library" /cell_line="HT1080" /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, In press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."		
FEATURES	Source		
BASE COUNT	93 a 139 c 95 g 81 t		
ORIGIN			
Query Match	48.9%; Score 350.8; DB 12; Length 409;		
Best Local Similarity	99.2%; Pred. No. 1.3e-71;		
Matches	363; Conservative 0; Mismatches 2; Indels 1; Gaps 1;		
Db	354	AAAGAGGCTGTATACAGAGGCGTGCGTGAACCTACAGGGTGCCACACTGAC-412	
Y	11	AAAAAAAAAGGCTGTATACAGAGGCGTGCGTGAACCTACAGGGTGCCACACTGAC-412	
Db	25	AAAAGGGCTGTATACAGAGGCGTGCGTGAACCTACAGGGTGCCACACTGAC-84	
Y	413	TTATGAGACCCCATGTGCATGTGGCATCAGCTCATAGGCAACTAATGATGGGTGCC-472	
Db	85	TTATGAGACCCCATGTGCATGTGGCATCAGCTCATAGGCAACTAATGATGGGTGCC-144	
Y	473	ACACCCAGGCGCATCGGGCAGCCAGGAGTCTACTGCGTGGTGCGTCCAGGAGCC-532	
Db	145	ACACCCAGGCGCATCGGGCAGCCAGGAGTCTACTGCGTGGTGCGTCCAGGAGCC-204	
Y	533	CTGAGGTCCAACTATGTGCTCAAAGACACCGGGATGTGCACGCTACACTCTCTCCAGGC-592	
Db	205	CTGAGGTCCAACTATGTGCTCAAAGACACCGGGATGTGCACGCTACACTCTCTCCAGGC-264	
Y	593	AACGAGCTACACACTATCGAAGTATGGGACACACTACCGGTCCCGCGAGAGCCCTGCT-652	
Db	265	AACGAGCTACACACTATCGAAGTATGGGACACACTACCGGTCCCGCGAGAGCCCTGCT-324	
Y	653	GATCGGACACCCATCTCTCCCTCCCATGGCCAAAGACCCCACTGTCTCTCTCAATA-712	
Db	325	GATCGGACACCCATCTCTCTCCCTCCCATGGCCAAAGACCCCACTGTCTCTCTCAATA-384	

OY	713	AAGATG 718
Db	385	AAGATG 390
RESULT 4		
LOCUS	AK008335	
DEFINITION	AK008335	682 bp mRNA linear HTC 19-JAN-2002
ACCESSION	AK008335	Mus musculus adult male small intestine cDNA, RIKEN full-length
VERSION	AK008335.1	enriched library, clone:2010107E11;peptidoglycan recognition
KEYWORDS	HTC; CAP trapper.	protein, full insert sequence.
SOURCE	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to	
ORGANISM	mRNA, clone:libRIKEN full-length enriched mouse cDNA library	
REFERENCE	Clone:2010107E11.	
ATTNORS	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
ATTNORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	Carninci,P. and Hayashizaki,Y.	
ATTNORS	High-efficiency full-length cDNA cloning	
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE	99279253	
PUBMED	10349636	
REFERENCE	2	
ATTNORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,	
REFERENCE	Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to	
MEDLINE	Prepare full-length cDNA libraries for rapid discovery of new genes	
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)	
REFERENCE	3	
ATTNORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,	
REFERENCE	Kono,H., Aiyama,Y., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M.,	
JOURNAL	Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,	
MEDLINE	Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiyagi,K.,	
PUBMED	Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ozuga,E., Watanabe,M.,	
REFERENCE	Toneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuno,S., Kawaj,J.,	
JOURNAL	Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.	
MEDLINE	RIKEN integrated sequence analysis (RISA) system--384-format	
PUBMED	sequencing pipeline with 384 multichannel sequencer	
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)	
ATTNORS	4	
REFERENCE	Kawai,J., Shingawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,	
JOURNAL	Arakawa,T., Hara,A., Fukunishi,Y., Kono,H., Adachi,J., Fukuda,S.,	
MEDLINE	Aizawa,K., Izawa,K., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,	
PUBMED	Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,	
REFERENCE	Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,	
JOURNAL	Frieschmann,W., Gaasterland,T., Gissi,C., King,B., Kochwa,H.,	
MEDLINE	Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,	
PUBMED	Quackenbush,J., Schriml,L.M., Staab,D.F., Suzuki,R., Tomita,M.,	
REFERENCE	Wagner,U., Mashio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,	
JOURNAL	Baldirelli,R., Barsh,G., Blake,J., Bottelli,D., Bojunga,N.,	
MEDLINE	Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bull,C.,	
PUBMED	Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,	
REFERENCE	Hofmann,L., Humm,D.A., Kamiya,M., Lee,N.H., Lyons,P.,	
JOURNAL	Marichioni,L., Mashima,J., Mazzarelli,J., Mommaerts,P., Nordone,P.,	
MEDLINE	Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,	
PUBMED	Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,	
REFERENCE	Toyo-oka,K., Wang,K.H., Weitz,C., Whitaker,C., Wilming,L.,	
JOURNAL	Wynshaw-Boris,A., Toshida,K., Hasegawa,Y., Kawaji,H., Kohsaki,S.	
MEDLINE	and Hayashizaki,Y.	
PUBMED	Functional annotation of a full-length mouse cDNA collection	
REFERENCE	Nature 409 (6821), 685-690 (2001)	
JOURNAL	5 (bases 1 to 682)	

BASE COUNT	163 a	214 c	179 g	147 t
ORIGIN				

Query Match 44.0%; Score 316; DB 13; Length 703;
Best Local Similarity 75.8%; Pred. No. 1,7e-63;
Matches 391; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
QY 135 CCTGCTGACGCCCATAGTCCCGGAGACAGTGAAGGCCCTGCGATCAGATGCGCCC 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88 CCGCTGAGTTTCATCGTGGCCCGAGTGAAGTGAAGGCCCGCCATCGAGTGCCTTA 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 AGCAGTGAAGCTGCTTACGCTATGCTGCTATGCGACAGCGGCGGAGCAGCTGCA 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 148 GCGGCTGGGGACCCAGTGTGCTGATCTACACACAGCGGAGCTTCGCA 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 ACACCCCCCTGCTGCGACAGAGCCCGGAAATGTCAGCACTACACATGAAGACAC 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 208 ACAGCCCGGACTCTCTGTGAACAGACAGGCCCGCATGTGACGCTTACACAGAAAGAGC 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 TGGGCTGTGTCAGCTGGGCTCAACTCTGATTTGAGAGAGAGCGGCTGTATACAGAG 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 268 TGGGCTGTGTCAGCTGTGCTCAACTCTTATTTGAGAGAGAGGCTCATGTCTATGAG 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 375 GCGGTGGCTGGAACCTTCACAGGGTGCCCATCAGTCTATTATGGAACCCCATGCTCAT 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 388 GCGGAGGCTGGAACATCAAGGGTGACACAGAGGCCCATCTGGAATCCCATGTATATG 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 435 GCATCAGCTTATGGGCACTACATGATGGGTGCGCCACACGCCCGCATCCGGGAG 494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 388 GCATCAGCTTATGGGCACTACATGATGGGTGCGCCACACGCCCGCATCCGGGAG 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 495 CCCAGGCTTACTGCGCTCGGTGTGCTGTCAGAGGACCCCTGAGGTCACATATGCTCA 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 448 CCCCTAACTCTTGGAATGTGGGGTGTCTCGGGGCTTCTTGAGATCCACTATGAGTCA 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 555 AAGGACACCGGATGAGCGCTACACTCTTCACAGGCAACACAGCTCTACACATGATCC 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 508 AAGGACACCGGATGAGCGCTACACTCTCTTCACAGGTCACCACTCTATCAAGTATCC 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 615 AGAATTGGCCACACTACCGCTCCCTCGAGGCCCTG 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 568 AAGCTGGGAACTACCAAGAGTGAAGAGACCTTG 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6 589 bp mRNA linear EST 06-FEB-2001
BG174272
LOCUS BG174272
DEFINITION 60233457JF1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4457680 5',
mRNA sequence.
ACCESSION BG174272
VERSION BG174272.1 GI:12680975
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 589)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov/
Plate: LMNL0254 row: n column: 17
High quality sequence stop: 570.
Location/Qualifiers

FEATURES
source 1..589
/organism="Mus musculus"
/strain="FVB/N"

/db_xref="taxon:10090"
/clone="IMAGE:4457680"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 126 a 174 c 161 g 128 t

ORIGIN

Query Match 43.9%; Score 315; DB 12; Length 589;
Best Local Similarity 77.0%; Pred. No. 2.9e-63;
Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```
QY 135 CCTGCTGACGCCCATAGTCCCGGAGACAGTGAAGGCCCTGCGATCAGATGCGCCC 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 50 CCGCTGAGTTTCATCGTGGCCCGAGTGAAGTGAAGGCCCGCCATCGAGTGCCTTA 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 AGCAGTGAAGCTGCTTACGCTATGCTGCTATGCGACAGCGGCGGAGCAGCTGCA 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 110 GCGGCTGGGGACCCAGTGTGCTGATCTTCACACAGACCGGAGCTTCTGCA 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 ACACCCCCCTGCTGCGACAGAGCCCGGAAATGTCAGCACTACACATGAAGACAC 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 170 ACAGCCCGGACTCTCTGTGAACAGACAGGCCCGCAATGTGACGATTCACACAGAAATGAGC 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 TGGGCTGTGTCAGCTGGGCTCAACTCTGATTTGAGAGAGAGCGGCTGTATACAGAG 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 230 TGGGCTGTGTCAGCTGTGCTCAACTCTTATTTGAGAGAGAGGTCATGTATGAG 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 375 GCGGTGGCTGGAACCTTCACAGGGTGCCCATCAGTCTATTATGGAACCCCATGCTCAT 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 290 GCGGAGGCTGGAACATCAAGGGTGACACAGAGGCCCATCTGGAATCCCATGTCTATG 349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 435 GCATCAGCTTATGGGCACTACATGATGGGTGCGCCACACGCCCGCATCCGGGAG 494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 350 GCATCAGCTTATGGGCACTACATGATGGGTGCGCCACACGCCCGCATCCGGGAG 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 495 CCCAGGCTTACTGCGCTCGGTGTGCTGTCAGAGGACCCCTGAGGTCACATGATGCTCA 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 CCTTAATCTTCTGGAATGTGGGGTGTCTCGGGGCTTCTTGAGATCCACTATGAAGTCA 469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 555 AAGGACACCGGATGAGCGCTACACTCTTCACAGGCAACACAGCTCTACACATGATCC 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 470 AAGGACACCGGATGAGCGCTACACTCTCTTCACAGGTCACCACTCTATCAAGTATCC 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 615 AGAATTGGCCACACTACCG 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 530 AAGCTGGGAACTACCAAGAGTGAAGAGACCTTG 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7 610 bp mRNA linear EST 05-JUL-2001
B1154844
LOCUS B1154844
DEFINITION 602902826P1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5032278 5',
mRNA sequence.
ACCESSION B1154844
VERSION B1154844.1 GI:14614845
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 610)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

DB 476 AAGGACACCGGGATGTGCAAGACACTCTCTCCAGGTGACCACTCTATCAGTATCC 535
 QY 615 AGAATTGGCCACACTACCG 633
 DB 536 AAGGCTGGGAGACACTACCG 554

RESULT 9
 B1155774
 LOCUS 602904343p1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5033867 5',
 DEFINITION mRNA sequence.
 ACCESSION B1155774
 VERSION B1155774.1 GI:14615775
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 630)
 NIH-MGC http://mgc.nci.nih.gov/.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM1094 row: f column: 12
 High quality sequence stop: 637.
 Location/Qualifiers
 1. 650
 /organism="Mus musculus"
 /strain="129,C57BL/6J,FVB/N"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:5033867"
 /clone_id="NCI_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /dev_stage="10 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
 Reference for transgenic model: Xu et al., Nature Genetics
 22, 37-43 (1999)."

BASE COUNT 137 a 199 c 179 g 135 t
 ORIGIN

Query Match 43.9%; Score 315; DB 13; Length 650;
 Best Local Similarity 77.0%; Pred. No. 3e-63;
 Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 135 CCTGCTGACGCCCATAGTGTGCGGAGAGAGAGGCGCTGAGTACAGATGCGGCC 194
 DB 114 CCTGCTGACGTTTATGTCGCGCGAGTAGTGAGAGGCGCTGAGTACAGATGCGGCC 173
 QY 195 AGACACGAGAGCGCTTACGATATGTGTGATGATGACAGAGGCGGAGAGAGTCA 254
 DB 174 GCGGCTGGGGACACGATGTCGATGATGATGATGATGATGATGATGATGATGATG 233
 QY 255 ACAACCCCGCTGTCGACAGAGAGGCGGAGTGTGACAGTACACATGGAAGAC 314
 DB 234 ACAGCCCGGAGTCTCTGTGAACAGAGAGGCGGAGTGTGACAGTACACATGGAAG 293
 QY 315 TGGGCTGTGTCGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 374
 DB 294 TGGGCTGTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 353

QY 375 GCGGTGCTGGAATTCACGGGTGCCCATGACGTCACTATATGGAACCCATGTCCATTTG 434
 DB 354 GCGAGGCTGTGGAACATCAAGGAGTACACACAGGCGCCATCTGGAATCCATGTCTATTG 413
 QY 435 GCATGACCTTCATGGGAGTACATGATGATGATGATGATGATGATGATGATGATGATG 494
 DB 414 GCATGACCTTCATGGGAGTACATGATGATGATGATGATGATGATGATGATGATGATG 473
 QY 495 CCCAGGCTTACGTGCGCTGCGGTGTGTCAGGAGAGCCCTGAGGTCCACTATGTGCTCA 554
 DB 474 CCTACATCTCTTGTGATGTGGGCTGTCTCGGGCTTCTTGACATCCACTATGAACTCA 533
 QY 555 AAGGACACCGGAGTGTGACAGCTCTCTCCAGGCAACAGCTTACACCTATCC 614
 DB 534 AAGGACACCGGAGTGTGCAAGGACACTCTCTCCAGGTACCAACTATCAGGTATCC 593
 QY 615 AGAATTGGCCACACTACCG 633
 DB 594 AAGGCTGGGAGACACTACCG 612

RESULT 10
 B1650838
 LOCUS 603297736p1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5338401 5',
 DEFINITION mRNA sequence.
 ACCESSION B1650838
 VERSION B1650838.1 GI:1565074
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 676)
 NIH-MGC http://mgc.nci.nih.gov/.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM11859 row: g column: 10
 High quality sequence stop: 676.
 Location/Qualifiers
 1. 676
 /organism="Mus musculus"
 /strain="129,C57BL/6J,FVB/N"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:5338401"
 /clone_id="NCI_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /dev_stage="10 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
 Reference for transgenic model: Xu et al., Nature Genetics
 22, 37-43 (1999)."

FEATURES
 source

BASE COUNT 156 a 204 c 174 g 142 t
 ORIGIN

Query Match 43.9%; Score 315; DB 13; Length 676;
 Best Local Similarity 77.0%; Pred. No. 3e-63;
 Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 135 CCTGCTGACGCCCATAGTGTGCGGAGAGAGAGGCGCTGAGTACAGATGCGGCC 194
 DB 114 CCTGCTGACGTTTATGTCGCGCGAGTAGTGAGAGGCGCTGAGTACAGATGCGGCC 173

Db 65 CCTCTGACATTCATGTCGCCCCGAGTAGAGAGGAGCCCTGCATCCGATGCTCTA 124
 Oy 195 AGACCTGAGGCTGCGCTTACGCTATGTGGTGGTATGGCAGACAGCGGCGAGCTGCA 254
 Db 125 GCGGCTGGGGCAGCCAGTTCGCTGCTGATCTCAGACACAGCGGCGAGCTTGTCA 184
 Oy 255 ACACCCCGGCTGCTGTCAGCAGAGCGCCGAGTGTGCAGCATTACACATGAAGAC 314
 Db 185 ACAGCCCGGAGCTCTCTGAGACAGAGCGCCGAGTGTGCAGCATTACACATGAAGAC 244
 Oy 315 TGGGCTGTGTCGAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAGG 374
 Db 245 TGGGCTGTGTCGAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAGG 304
 Oy 375 GCGGTGGCTGGAAGTGGAGGTCGCGACATGATCTATGAGACCCCATGTCATTG 434
 Db 305 GCGGAGGCTGGAAGTGGAGGTCGCGACATGATCTATGAGACCCCATGTCATTG 364
 Oy 435 GCATCAGCTTCATGAGGACATGATGATGGTGGTGGCCAGACCCAGCCATCCGGGAG 494
 Db 365 GCATCAGCTTCATGAGGACATGATGATGGTGGTGGCCAGACCCAGCCATCCGGGAG 424
 Oy 495 CCCAGGCTTACTGCGCTGCGGTGTGCTCAGGAGCGCTGAGGTCCACTATGTCTCA 554
 Db 425 CCTTAATCTTGTGGAATGTGGGTGTCTCGGGCTTCTTGAGATCCAACTATGAAATCA 484
 Oy 555 AAGGACACCGGAGTGTGACAGCTACACTCTCTCCAGGACACGAGCTTACCACTCATCC 614
 Db 485 AAGGACACCGGAGTGTGACAGCTACACTCTCTCCAGGACACGAGCTTATCATGCTATCC 544
 Oy 615 AGAATTGGCCACACTACCG 633
 Db 545 AAAGCTGGGAACACTACCG 563

RESULT 11

BQ951956

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BQ951956 686 bp mRNA linear EST 21-AUG-2002
 AGENCOURT 8951306 NCI_CGAP_Co24 Mus musculus cDNA clone
 IMAGE:6474853 5', mRNA sequence.
 BQ951956
 BQ951956.1 GI:22367434
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 686)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bcr-remail.nih.gov
 Tissue Procurement: The Cepho Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M14012 row: k column: 14
 High quality sequence stop: 589.
 Location/Qualifiers

FEATURES

SOURCE

1. 686

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone_image="IMAGE:6474853"

/clone_lib="NCI_CGAP_Co24"

/lab_host="DH10B (TI phase-resistant)"

/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
 Site:2: SalI; Cloned: unidirectionally. Primer: Oligo dT.
 Average insert size 1.6 Kb. Constructed by Life

BASE COUNT 157 a 207 c 172 g 141 t 9 others
 ORIGIN
 Query Match. 43.9% Score 315; DB 14; Length 686;
 Best Local Similarity 77.0%; Pred. No. 3e-63;
 Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Oy 135 CCTGCTGACGCCCATAGTCCCGGAGAGAGTGGAGAGCCCTGATCAGATGCGGCC 194
 Db 63 CCTGCTGAGTTGATGCTGTCGCGCCGAGTAGTGGAGGCGCCCTGCAATCCAGTCTCTA 122
 Oy 195 AGACCTGAGGCTCCCTTACGCTATGTGTGTATGTCACACAGCGGCGAGCAGCTGCA 254
 Db 123 GCGGCTGGGGCAGCCAGTTCGCTGATGATCTCAGACAGCGGCGAGCTTCTGCA 182
 Oy 255 ACACCCCGGCTGCTGTCAGCAGAGCGCCGAGTGTGCAGCATTACACATGAAGACAC 314
 Db 183 ACAGCCCGGAGCTCTCTGAGACAGAGCGCCGAGTGTGCAGCATTACACATGAAGAC 242
 Oy 315 TGGGCTGTGTCGAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAGG 374
 Db 245 TGGGCTGTGTCGAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAGG 302
 Oy 375 GCGGTGGCTGGAAGTGGAGGTCGCGACATGATCTATGAGACCCCATGTCATTG 434
 Db 305 GCGGAGGCTGGAAGTGGAGGTCGCGACATGATCTATGAGACCCCATGTCATTG 362
 Oy 435 GCATCAGCTTCATGAGGACATGATGATGGTGGTGGCCAGACCCAGCCATCCGGGAG 494
 Db 365 GCATCAGCTTCATGAGGACATGATGATGGTGGTGGCCAGACCCAGCCATCCGGGAG 422
 Oy 495 CCCAGGCTTACTGCGCTGCGGTGTGCTCAGGAGCGCTGAGGTCCACTATGTCTCA 554
 Db 425 CCTTAATCTTGTGGAATGTGGGTGTCTCGGGCTTCTTGAGATCCAACTATGAAATCA 482
 Oy 555 AAGGACACCGGAGTGTGACAGCTACACTCTCTCCAGGACACGAGCTTACCACTCATCC 614
 Db 485 AAGGACACCGGAGTGTGACAGCTACACTCTCTCCAGGACACGAGCTTATCATGCTATCC 542
 Oy 615 AGAATTGGCCACACTACCG 633
 Db 543 AAAGCTGGGAACACTACCG 561

RESULT 12

BG244455

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG244455 688 bp mRNA linear EST 13-FEB-2001
 602356590F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4485071 5',
 mRNA sequence.
 BG244455
 BG244455.1 GI:12754270
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 688)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bcr-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M10326 row: c column: 24
 High quality sequence stop: 587.
 Location/Qualifiers

FEATURES

Query Match	43.8%	Score 314.6	DB 13	Length 842
Best Local Similarity	76.8%	Pred. No. 3.8e-63		
Matches 383	Conservative 0	Mismatches 116	Indels 0	Gaps 0
DB	135	CCTGCTGAGGCCCATATGTCGCCCGGAGAACGATGGAAAGGCCCTGCGATCAGATGCGGCC	194	
DB	34	CCTCTGCGAGTTTACATCTGCTGCTGCGCGAGTGTGGAGGCCCTCGCATCCGAGGCTCTA	93	
DB	195	AGCACTGAGCCCTGCTTACGCTATGTGTGTATGCGACACGCGGCGAGCAGCTGCA	254	
DB	94	GCGCGCTGGGCGACCAATGCTACGTACGTACGTACGTACGTACGTACGTACGTACGTAC	153	
DB	255	ACACCCCGCCCTGCTGTCGACAGCAGGCGCGGAAATGTGACACTTACCATGAGACAC	314	
DB	154	ACAGCGCGGATCTCTGTAAACAGCAGCGCGCATGTGACACTTACCATGAGATGAGC	213	
DB	315	TGGGCTGTGCGAGCTGGGCTGACAACTCTCGATGTGGAGAAAGAGGGGCTGTATACGAG	374	
DB	214	TGGGCTGTGCGAGCTGGGCTGACAACTCTCTATGTGGAGAGAGCGGTCTATATGAG	273	
DB	375	GCGCTGTGCGAGCTTACAGGCGTCCCATCTCAGCTCACTTATGAGAACCCCATGTGCA	434	
DB	274	GCGAGGCTGGAAATCAGGCGTACCAACAGGCGCGCATCTGGAATCCATGCTCTANTG	333	
DB	435	GCATCAGCTTATGAGGCACTACATGATGAGGCGCGCACACCCGAGCCATCCGGGAG	494	
DB	334	GCATCAGCTTATGAGGCACTTACATGAGCGGATACCCGAGAAAGCGGCGCTCCGCTG	393	
DB	495	CCGAGGCTTACTGCGCTGCGGTGTGCTCAGGAGCGCCCTGAGGTCCAACTATGTGCTA	554	
DB	394	CCCTAAATCTTCTGGAAATGTTGGGTGTCTGCGGCTCTCTGAGATCCAACTATGAACTCA	453	
DB	555	AAGGACACCGGAGTGCAGAGCTACATCTCTCCAGGCAACAGCTTACACCTCATCC	614	
DB	454	AAGGACACCGGAGTGCAGAGCTCTCTCCAGGTCACAACTTATCAGGTATCC	513	
DB	615	AGAAATGGCACACTACCG 633		

```

BASE COUNT      130 a      216 c      156 g      124 t      1 others
ORIGIN
/organism="Brugia malayi"
/strain="FRS Labs"
/db_xref="taxon:6279"
/clone="SMFCA1124"
/clone_1fb="Brugia malayi microfilaria cDNA (SAW94LS-Bmf
)
/lab_host="X1-Blue MRF"
/note="Vector: lambda Unizap XR; Site.1: EcoR I; Site.2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from microfilariae of Brugia malayi
isolated from birds and converted to double stranded cDNA
using reverse transcriptase and oligo(dT) followed by
Rnase H and Dnasep I. The library had 3.5 x 10E5
independent recombinants and average insert size was 900
base pairs. The library was constructed by Lori Saunders.
The library is available from Dr. S.A. Williams, email
genomes@smth.edu."

```

RESULT	15
BG975104	
LOCUS	
DEFINITION	BG975104 645 bp mRNA linear EST 12-JUN-2001
ACCESSION	F0284336F1 NCI-CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979057 5'
VERSION	BG975104
KEYWORDS	BG975104.1 GI:14362741
SOURCE	EST.
ORGANISM	house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 645) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D.
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
42.7%;	Score 306.6;	DB 13;	Length 645;			
77.3%;	Pred. No. 2.7e-61;					
0;				109;	Indels	0; Gaps 0;

BASE COUNT	ORIGIN
134 a	202 c 176 g 133 t

135	109	109
CTCTCTGCAGCCCAATGATGTGCCCCGGGAACGAGTGGAAGGCCCTTGATCGATCAGAGTGGCCCC	CTCTCTCTGCAGCCCAATGATGTGCCCCGGGAACGAGTGGAAGGCCCTTGATCGATCAGAGTGGCCCC	CTCTCTCTGCAGCCCAATGATGTGCCCCGGGAACGAGTGGAAGGCCCTTGATCGATCAGAGTGGCCCC
168	168	168

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 00:03:09 ; Search time 2720 seconds

(without alignments)
7682.285 Million cell updates/sec

Title: US-09-462-625-3

Perfect score: 718

Sequence: 1 ctgagttactgagccacag.....ctcctctccaataagatg 718

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_da: *
2: gb_hg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_or: *
21: em_ov: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hg_hum: *
31: em_hg_inv: *
32: em_hg_other: *
33: em_hg_mus: *
34: em_hg_pla: *
35: em_hg_rod: *
36: em_hg_mam: *
37: em_hg_vit: *
38: em_sy: *
39: em_hgo_hum: *
40: em_hgo_mus: *
41: em_hgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	673	93.7	724	9	AF242517	AF242517 Homo sapi
2	653.2	91.0	690	9	AF076483	AF076483 Homo sapi
3	393.6	54.8	700	4	CDR131676	AJ131676 Camelus d
4	349.4	48.7	688	4	AY083309	AY083309 Bos tauru
5	316.6	44.1	549	6	AR124884	AR124884 Sequence
6	316.6	44.1	669	10	AF193843	AF193843 Mus muscu
7	316.6	44.1	680	10	AF076482	AF076482 Mus muscu
8	315	43.9	713	10	BC005582	BC005582 Mus muscu
9	302.6	42.1	678	10	MMRNATMS1	X86374 M.musculus
10	299	41.6	166500	9	AC007785	AC007785 Homo sapi
11	295.6	41.2	630	10	AF154114	AF154114 Rattus no
12	185.4	25.8	259	4	SSC310355	AJ10355 Sus scrof
13	161.4	22.5	5358	4	CDR409286	AJ409286 Camelus d
14	147.4	20.5	166867	3	AC007085	AC007085 Drosophi
15	147.4	20.5	169509	3	AC007303	AC007303 Drosophi
16	147.4	20.5	172838	2	AC020451	AC020451 Drosophi
17	147.4	20.5	259336	3	AE003836	AE003836 Drosophi
18	145	20.2	651	3	AY060645	AY060645 Drosophi
19	145	20.2	652	3	AF207542	AF207542 Drosophi
20	136.4	19.0	717	3	AF207541	AF207541 Drosophi
21	136.4	19.0	1162	3	AY075293	AY075293 Drosophi
22	132.2	18.4	1107	6	AX119920	AX119920 Sequence
23	132.2	18.4	1125	6	AX119917	AX119917 Sequence
24	127	17.7	4205	10	MMTAG7	Y12088 Mus muscu
25	122	17.0	182897	2	AC110846	AC110846 Rattus no
26	117.6	16.4	1110	6	AX119918	AX119918 Sequence
27	117.6	16.4	1128	6	AX119915	AX119915 Sequence
28	117.6	16.4	1194	9	AY035377	AY035377 Homo sapi
29	117.4	16.4	1858	9	AF242518	AF242518 Homo sapi
30	115.8	16.1	1173	9	AY035376	AY035376 Homo sapi
31	97.8	13.6	521	6	E16478	E16478 Bombyx mori
32	97.8	13.6	590	6	E16479	E16479 Bombyx mori
33	97.8	13.6	753	6	E16480	E16480 Bombyx mori
34	96.2	13.4	735	6	AB016249	AB016249 Bombyx mo
35	94.6	13.2	182897	2	AC110846	AC110846 Rattus no
36	94	13.1	1794	9	AF384856	AF384856 Homo sapi
37	93.2	13.0	718	3	AF441723	AF441723 Bombyx mo
38	92.4	12.9	1374	3	AY058258	AY058258 Drosophi
39	92.4	12.9	1375	3	AF313391	AF313391 Drosophi
40	92.4	12.9	2751	9	AK055882	AK055882 Homo sapi
41	90.2	12.6	1369	3	AB084190	AB084190 Drosophi
42	89	12.4	670	3	AF076481	AF076481 Trichoplu
43	88.8	12.4	622	3	AB017520	AB017520 Bombyx mo
44	87	12.1	55863	2	AC018100	AC018100 Drosophi
45	87	12.1	157225	3	AC010042	AC010042 Drosophi

ALIGNMENTS

RESULT 1
AF242517
LOCUS AF242517 724 bp mRNA linear PRI 15-NOV-2000
DEFINITION Homo sapiens hypothetical protein SBB168 mRNA, complete cds.
ACCESSION AF242517
VERSION AF242517.1 GI:9802032
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 724)
AUTHORS Wan, T., Zhang, W. and Cao, X.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Department of Immunology, Second Military

|||||
 Db 301 AGACATGGGCTGTGCGACGTGGCTACACTCTCTGATGTGAGAGACGGGCTGTAT 360
 369 AGGAGGGCGGTGGTGAACCTCAGGGGTGCCACCTCAGTACTATATGAACCCATAT 428
 Db 361 ACGAGGGCGGTGGTGAACCTCAGGGGTGCCACCTCAGTACTATATGAACCCATAT 420
 429 CCATTGGCATGAGTCTTATGGGCACTACATGATCGGGTGGCCACACCCAGGCCATCC 488
 Db 421 CCATTGGCATGAGTCTTATGGGCACTACATGATCGGGTGGCCACACCCAGGCCATCC 480
 489 GGGCAGCCAGGGTCTACTGCGCTCGGTGGTGTGAGGAGAGCCCTGAGGTCCAACTATG 548
 Db 481 GGGCAGCCAGGGTCTACTGCGCTCGGTGGTGTGAGGAGAGCCCTGAGGTCCAACTATG 540
 549 TGCTAAAGAGACAGGGGATGTGAGCGGTACACTCTCTCAGGCAACCAAGCTCTACAC 608
 Db 541 TGCTAAAGAGACAGGGGATGTGAGCGGTACACTCTCTCAGGCAACCAAGCTCTACAC 600
 609 TCATCAGAAATTTGGCCACACTACCGCTCCCTGAGGCGCTGTGATCCGACCCCATTC 668
 Db 601 TCATCAGAAATTTGGCCACACTACCGCTCCCTGAGGCGCTGTGATCCGACCCCATTC 660
 669 CTCCCTCCCATGGCCAAAACCCCATGT 698
 Db 661 CTCCCTCCCATGGCCAAAACCCCATGT 690

RESULT 3
 CDR131676 700 bp mRNA linear NAM 21-DEC-2000
 LOCUS Camelus dromedarius mRNA for peptidoglycan recognition protein.
 DEFINITION AJ131676.1 GI:11990123
 VERSION peptidoglycan recognition protein.
 KEYWORDS Arabid camel.
 SOURCE Camelus dromedarius
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 REFERENCE 1 (bases 1 to 700)
 AUTHORS Kappeler S.R., Farah Z. and Puhan Z.
 TITLE Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan Recognition Protein
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 700)
 AUTHORS Kappeler S.R.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1998) Kappeler S.R., Institute of Food Science, Swiss Federal Institute of Technology, LFO F26, 8092 Zurich, SWITZERLAND

FEATURES
 source location/Qualifiers
 1..700
 /organism="Camelus dromedarius"
 /sub_species="Somali breed"
 /db_xref="taxon:9838"
 /tissue_type="lactating mammary gland"
 25..606
 /codon_start=1
 /product="peptidoglycan recognition protein"
 /protein_id="CAC19553.1"
 /db_xref="GI:11990123"
 /translation="MTRHCVALWALLALSLGAARDPPACGSIIVRRWRALASEC
 REHLTRPVVVVSHRSHCDPASCQAQAOAYQSYHRLGCGDVGNFLIGEDGI
 VVEGRGNINIGAHAGPTWNTISIGISFGNVMNPPPALRAAQNLLACVALGALR
 SNVEYGHDRVOPFLSPGDRLEIITQWSHRA"
 25..87
 sig_peptide
 88..603
 mat_peptide
 /product="peptidoglycan recognition protein"
 678..683
 poly_a_signal
 130 a 247 c 201 g 122 t
 BASE COUNT
 ORIGIN
 Query Match 54.88; Score 393.6; DB 4; Length 700;

Best Local Similarity 77.1%; Pred. No. 1,3e-77;
 Matches 507; Conservative 0; Mismatches 144; Indels 7; Gaps 2;
 51 CTATGTCCGCGGCTCTATGCTGCTTGTGGGCTCTCCACCTCTTGATCGATGGAG 110
 Db 23 CCATGACCGGCGACTCGTGTCTCTGCTGGGCTCTCCCTCGGCTCTCAGCTGGAG 82
 111 CGGCTCAGAGACAGAAAGACCGGCTGTGCGAGCCCATAGTCCCGGAGACAGTGA 170
 Db 83 CGGCTG-----AGAAAGACCGGCGGCTGTGCGGCTCTCATCGGCCCGCGAGAGTGA 136
 171 AGGCGCTGCAATCAGATGAGTGGCGCCAGCACACCTGAGCCCTAGCTATGTGGTAT 230
 Db 137 GGGCGCTGGCGCTCCGATGTCAGAGAAAGCTTAACACGCGGCTTACGTGTGGTGT 196
 231 CGACACGGCGGGGACAGACCTGCAACACCGCGCTGTGCCAGCAGAGCCCGAATG 290
 Db 197 CGCACACTGCGGGGACGACCTGCGACACCGCGGCTTGTGCGCGCAGAGCCCGAGAG 256
 291 TGCAGACATACCAATGAAAGACACTGGGCTGTGCGACGTGGGCTTCACTTCTGATTG 350
 Db 257 TGCAAAAGTACATGTGCGGAACCTGGGCTGTGCGACGTGGGCTTCACTTCTGATTG 316
 351 GAGAAAGACGGGCTGTATACAGAGGCGGTGGCTGGAACCTTACAGGGGTGCCACTCAG 410
 Db 317 GAGAAAGATGGGCTCGTGTACGAAAGCGGAGCTGGAACATCAAGGGCGCCACGAGTC 376
 411 ACTTATGGAACCCCATGTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 470
 Db 377 CCACCTGGAAACCCCATATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 436
 471 CCACACCCAGGCGCAATTCGGGAGCGCCAGGCTTACTGGCTGCGGTGTGGCTCAGGAG 530
 Db 437 CCGCGGCGCGGCGCTCGGCGGAGCCCAAAATCTGTGCTGTGGCTGTGGCTGTGGAG 496
 531 CCGTGAAGTCACTATGTGCTCAAAAGACACCGGATGTGACGCTGACACTCTCCAG 590
 Db 497 CCGTGAAGTCACTATGTGCTCAAAAGACACCGGATGTGACGCTGACACTCTCCAG 556
 591 GCAACAGCTCTACCACTCATCAGAAATTTGGCCACACTACCGCTCCCTGAGGCCCTG 650
 Db 557 GTGACCGGCTCTACCAAAATTCATTCATTCATTCATTCATTCATTCATTCATTC 615
 651 GTGATCCGACCCCAATTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 708
 Db 616 CGGCTGGGACACCGCTCCATCCCATCTCTGTCAAAACCCCATCTGCTTCCCGC 673

RESULT 4
 AY083309 688 bp mRNA linear NAM 29-MAY-2002
 LOCUS Bos taurus oligosaccharide-binding protein mRNA, complete cds.
 DEFINITION AY083309.1 GI:19550241
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Bos taurus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 688)
 AUTHORS Tydell, C.C., Yount, N., Tran, D., Yuan, J. and Seilsted, M.E.
 TITLE Isolation, characterization, and antimicrobial properties of bovine oligosaccharide-binding protein. A microbicidal granule protein of eosinophils and neutrophils
 JOURNAL J. Biol. Chem. 277 (22), 19558-19664 (2002)
 MEDLINE 11880375
 PUBMED 122028028
 2 (bases 1 to 688)
 REFERENCE Yount, N.Y., Yuan, J., Tydell, C.C. and Seilsted, M.E.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAY-2002) Pathology, UC Irvine, 440D Medical Sciences 1, Irvine, CA 92697, USA

FEATURES
source
Location/Qualifiers
1..688
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_type="peripheral white blood cells"
29..601
/note="microbicide; undergoes N-terminal glutamine cyclization; peptidoglycan recognition protein; PRP"
/odon_start=1
/product="oligosaccharide-binding protein"
/protein_id="AAL87002.1"
/db_xref="GI:19550242"
/translation="MSRRYPLAWLILGLAODCSIVSRGKMGALASCSORL RQPVRYVVSHTGASVCTPASCORQAOVYHVERGMDVGFLLGEGDLVEG RGMRTGASGSPMNPRIAGISFPMGNMHRVPPASALRAQSLIACGAANGYLPNTE VGHKRDVQOQLSPGDELKXIQDPHFRV"
misc_feature
order(98..100,470..472)
/note="disulfide bond"
misc_feature
order(146..148,281..283)
/note="disulfide bond"
misc_feature
order(209..211,227..229)
/note="disulfide bond"
BASE COUNT 126 a 234 c 201 g 127 t
ORIGIN

Query Match 48.7%; Score 349.4; DB 4; Length 688;
Best Local Similarity 72.9%; Pred. No. 9.4e-68;
Matches 485; Conservative 0; Mismatches 161; Indels 19; Gaps 2;

51 CTATGTCGCCGCTCTATGCTGCTGCGGCTCTCCCGAGCTCTCGACTCGAG 110
111 CGGCTGAGGAGACAGAGACCGCGCTGCTGCAAGCCCAATAGTCCCGGAGACAGTGA 170
87 CGGCTCAAG-----CTGGGACACATCTGTCCTCCGGGAAAGTGGG 128
171 AGGCGCTGAGTACAGATGCGCCGACGACCTGACCTGCTATGCTGATG 230
129 GCGCCCTGGGATCCAACTGACAGGAGGCTGACAGCCCTGCTGCTGCTGCT 188
231 CGCACAGCGCGGAGAGCTGCAACACCCCGCTGCTGCGACAGAGCGGAGATG 290
189 CGCACAGCGCGGAGAGCTGCAACACCTCGGCTGCTGCGACAGAGCGGAGATG 248
291 TGGAGACTACCAACAAACACATGCGCTGCTGCGACAGAGCGGAGATG 350
249 TGGAGACTACCAACAGCGGAGCGGCTGCTGCGACAGAGCGGAGATG 308
351 GAGAGAGCGGCTGCTGATGAGAGCGGCTGCTGCACTGCAAGCTGCGACAGGTC 410
309 GAGAGAGCGGCTGCTGATGAGAGCGGCTGCACTGCAAGCTGCTGCTGCTG 368
411 ACTTATGGAACCCCATGCTGATGAGTGCATGCTGCACTGCAATGATGAGTGC 470
369 CCAGTGGGAACCCCATGCTGATGAGTGCATGCTGCACTGCAATGATGAGTGC 428
471 CCACACCCCGGAGCGGAGCGGCTGCTGCGACAGAGCGGAGTGCATGAGGAG 530
429 CCGCGGCTGCTGCTGAGGCGGCGGAGTGTGCTGCTGAGGCGGAGTGCATG 488
531 CCCTGAGGCTCAATATGCTGCTGCAAGAGACCGGATGCTGCACTGCTGCTG 590
489 ACCTGAGTCTCAATGAGTGCATGAGGAGACCGGATGCTGCACTGCTGCTG 548
591 GCAACAGCTCTACACCTGATCCAGATTTGCGCACACTACCGCTGCCCTGAGGCC 650
549 GGGAGGAGCTCTATTAATATCAGAGTGGCGGCACTACCGCGGAGTGGAGGCC 608
651 CTGATCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 710
609 TCGGCTTTC-TCAACACCCCATGCTGCAAGAGCGGCGGCTGCTGCTGCTGCT 667

QY 711 TAAAG 715
DB 668 TAAAG 672

RESULT 5
LOCUS ARI24884 549 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6172211.
ACCESSION ARI24884
VERSION ARI24884.1 GI:14110245
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 549)
AUTHORS Georgiev,G.P., Kiselev,S.L., Prokhorchouk,E.B. and Ostermann,E.
TITLE Nucleic acid encoding tag7 polypeptide
JOURNAL Patent: US 6172211-A 1 09-JAN-2001;
FEATURES
source Location/Qualifiers
1..549
/organism="unknown"
BASE COUNT 117 a 161 c 151 g 120 t
ORIGIN

Query Match 44.1%; Score 316.6; DB 6; Length 549;
Best Local Similarity 77.2%; Pred. No. 2e-60;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

135 CCTCTGACAGCCCATAGTGCCCGGAGAGTGAAGAGCCCTGACATGAGTGGCC 194
44 CTTCTGAGATTTCATGCTGCTGCGGAGTGAAGAGGCGCTGCTGCTGCTGCT 103
195 AGCACTGAGCTCTCCCTTACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 254
104 GCGCCCTGCGGCGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 163
255 ACACCCCGGCTGCTGCGACAGAGAGCGGAGTGTGCTGCTGCTGCTGCTGCTG 314
164 ACAGCCCGGACTCTGCTGACAGAGCGGCGGAGTGTGCTGCTGCTGCTGCTG 223
315 TGGGCTGCTGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 374
224 TGGGCTGCTGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 283
375 GCGGCTGCTGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 434
284 GCCAGGCTGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 343
435 GCATCAGCTTATGAGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 494
344 GCATCAGCTTATGAGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
495 CCCAGGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 554
404 CCCAATCTCTGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 463
555 AAGACACCGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 614
464 AAGACACCGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 523
615 AGAATGGGCACTACCG 633
524 AAGCTGGAGACCTACCG 542

RESULT 6
LOCUS AF193843 669 bp mRNA linear ROD 07-NOV-1999
DEFINITION Mus musculus Tag7-like protein mRNA, complete cds.
ACCESSION AF193843
VERSION AF193843.1 GI:6273360
KEYWORDS

SOURCE
ORGANISM Mus musculus.
REFERENCE Mus musculus.
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 669)
JOURNAL Slayton, W.B., Rigaa, A., Hancock, J.D., Zaugg, J.K., Le, T.V.,
REFERENCE Treutman, M.S., Spangrude, G.J., Carroll, W.L. and Schibler, K.R.
AUTHORS Granulocyte-colony stimulating factor up-regulates expression of
TITLE murine tag7 during myeloid differentiation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 669)
AUTHORS Slayton, W.B., Rigaa, A., Hancock, J.D., Zaugg, J.K., Le, T.V.,
TITLE Treutman, M.S., Spangrude, G.J., Carroll, W.L. and Schibler, K.R.
JOURNAL Direct Submission
FEATURES Submitted (11-OCT-1999) Pediatrics, University of Utah, 50 North
SOURCE Medical Drive, Room 2A126, Salt Lake City, UT 84132, USA
Location/Qualifiers
1. 669
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="32D cl3"
/cell_type="myeloblast"
25. 573
/codon_start=1
/product="TAG7-like protein"
/protein_id="AA06335.1"
/db_xref="GI:6273361"
/translation="MLPACALALALGATSCSFIVPSEMRALPSESSRLGHPVRY
VISHTAGSFNSPDSCEQOARVVOHYHKNELGMDVAVNPLIGEDGHYERGVNLIK
DHRTGPIWPMISGITGFPMGNFMDRVPARKRALNALNLECGVSRGFLRSNYEVKGRHDY
OSTLSPDQLYOVISWEHYRE"
BASE COUNT 144 a 212 c 173 g 140 t
ORIGIN

Query Match 44.1%; Score 316.6; DB 10; Length 669;
Best Local Similarity 77.2%; Pred. No. 1.9e-60;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

135 CCTGCTGAGCCCATGATGCGCCGAGACAGTGAAGGACCCCTGCGATCAGAGTGGCCC 194
Db CCTCTGACGTTTCACTCTGCGCCGAGTGAAGGACCCCTGCGATCAGAGTGGCCC 127
195 AGACCTGAGGCTGCGCTTACGCTATGTGTATCGACAGGGGGGACAGCTGCA 254
Db GCGGCTGGGGGACAGCTGCTACGTGTATCGACAGGGGGGACAGCTGCTGCA 187
255 ACACCCCGGCTGCTGCGACAGAGGCGCGAATGTGCAGCATACATGAAGACAC 314
Db ACACCCCGGCTGCTGCGACAGAGGCGCGAATGTGCAGCATACATGAAGACAC 247
315 TGGGCTGTGAGCTGGGGCTGACAACTTCGATGTGAGAGAGCGGCTCGATACGAG 374
Db TGGGCTGTGAGCTGGGGCTGACAACTTCGATGTGAGAGAGCGGCTCGATACGAG 307
375 GCGGCTGTGAGCTGAGGCTGCGGACAGTCACTTATGAAACCCATGTCATTG 434
Db GCGGCTGTGAGCTGAGGCTGCGGACAGTCACTTATGAAACCCATGTCATTG 367
435 GCATCAGCTTATGAGGCACTACATGATGGGTGCGCACACCCAGCCATCGGGGAG 494
Db GCATCAGCTTATGAGGCACTACATGATGGGTGCGCACACCCAGCCATCGGGGAG 427
495 CCCAGGCTACTGAGGCTGAGGCTGAGGAGCCCTGAGGCTGAGGCTGAGGCTGAG 554
Db CCCAGGCTACTGAGGCTGAGGCTGAGGAGCCCTGAGGCTGAGGCTGAGGCTGAG 487
555 AAGGACACCGGAGTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 614
Db AAGGACACCGGAGTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 488
615 AGAATTTGGCCACACTACCG 633
Db AAGGCTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 548
548 AAGGCTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 566

RESULT 7
AF076482 680 bp mRNA linear ROD 15-AUG-1998
LOCUS AF076482
DEFINITION Mus musculus peptidoglycan recognition protein precursor (Pgip)
ACCESSION AF076482
VERSION AF076482.1 GI:3342530
KEYWORDS
SOURCE
ORGANISM Mus musculus.
REFERENCE Mus musculus.
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 680)
AUTHORS Kang, D., Liu, G., Lundstrom, A., Gellus, E. and Steiner, H.
TITLE A peptidoglycan recognition protein in innate immunity conserved
JOURNAL from insects to humans
REFERENCE Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
MEDLINE 98374308
PUBMED 9707603
2 (bases 1 to 680)
REFERENCE Kang, D., Liu, G., Lundstrom, A., Gellus, E. and Steiner, H.
AUTHORS Direct Submission
TITLE Submitted (08-JUL-1998) Microbiology, Stockholm University,
JOURNAL Stockholm S-106 91, Sweden
Location/Qualifiers
1. 680
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="spleen"
1. 680
/gene="Pgip"
36. 584
/gene="Pgip"
/note="Pgip"
/codon_start=1
/product="peptidoglycan recognition protein precursor"
/protein_id="AAC31821.1"
/db_xref="GI:3342531"
/translation="MLPACALALALGATSCSFIVPSEMRALPSESSRLGHPVRY
VISHTAGSFNSPDSCEQOARVVOHYHKNELGMDVAVNPLIGEDGHYERGVNLIK
DHRTGPIWPMISGITGFPMGNFMDRVPARKRALNALNLECGVSRGFLRSNYEVKGRHDY
OSTLSPDQLYOVISWEHYRE"
BASE COUNT 146 a 216 c 174 g 144 t
ORIGIN

Query Match 44.1%; Score 316.6; DB 10; Length 680;
Best Local Similarity 77.2%; Pred. No. 1.9e-60;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

135 CCTGCTGAGCCCATGATGCGCCGAGACAGTGAAGGACCCCTGCGATCAGAGTGGCCC 194
Db CCTCTGACGTTTCACTCTGCGCCGAGTGAAGGACCCCTGCGATCAGAGTGGCCC 127
195 AGACCTGAGGCTGCGCTTACGCTATGTGTATCGACAGGGGGGACAGCTGCA 254
Db GCGGCTGGGGGACAGCTGCTACGTGTATCGACAGGGGGGACAGCTGCTGCA 187
255 ACACCCCGGCTGCTGCGACAGAGGCGCGAATGTGCAGCATACATGAAGACAC 314
Db ACACCCCGGCTGCTGCGACAGAGGCGCGAATGTGCAGCATACATGAAGACAC 247
315 TGGGCTGTGAGCTGGGGCTGACAACTTCGATGTGAGAGAGCGGCTCGATACGAG 374
Db TGGGCTGTGAGCTGGGGCTGACAACTTCGATGTGAGAGAGCGGCTCGATACGAG 307
375 GCGGCTGTGAGCTGAGGCTGCGGACAGTCACTTATGAAACCCATGTCATTG 434
Db GCGGCTGTGAGCTGAGGCTGCGGACAGTCACTTATGAAACCCATGTCATTG 367
435 GCATCAGCTTATGAGGCACTACATGATGGGTGCGCACACCCAGCCATCGGGGAG 494
Db GCATCAGCTTATGAGGCACTACATGATGGGTGCGCACACCCAGCCATCGGGGAG 427
495 CCCAGGCTACTGAGGCTGAGGCTGAGGAGCCCTGAGGCTGAGGCTGAGGCTGAG 554
Db CCCAGGCTACTGAGGCTGAGGCTGAGGAGCCCTGAGGCTGAGGCTGAGGCTGAG 487
555 AAGGACACCGGAGTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 614
Db AAGGACACCGGAGTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 488
615 AGAATTTGGCCACACTACCG 633
Db AAGGCTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 548
548 AAGGCTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 566

```

Db 319 GCCAGGCTGGACATCAAGGTGACACAGAGGCCCATCTGGAATCCCATGTATTG 378
QY 435 GCATCAGCTTATGGGCACTACATGATCGGGTCCCAACACCCAGCCATCCGGGAG 494
Db 379 GCATCAGCTTATGGGCACTACATGATCGGGTCCCAACACCCAGCCATCCGGGAG 438
QY 495 CCCAGGCTTACTGGCCGTGTGTGCTCAGAGGAGCCCTGAGTCAACTATGTCTCA 554
Db 439 CCCAATCTTCTGGAATGTGGGTGTCTCGGGCTCTCTGATGATCAATGAATCA 498
QY 555 AAGCAGCCGGGATGTGACGGGTACACTCTCTCCAGGCAACCGAGCTTACCACTATCC 614
Db 499 AAGCAGCCGGGATGTGACGGGTACACTCTCTCTCCAGGCAACCGAGCTTACCACTATCC 558
QY 615 AAGATGGCCACACTACCG 633
Db 559 AAGCTGGGACACTACCG 577

```

RESULT 8
LOCUS BC005582
DEFINITION Mus musculus, peptidoglycan recognition protein, clone MGC:11430
ACCESSION BC005582
VERSION BC005582.1 GI:13542755
KEYWORDS MGC.
SOURCE Mus mouse.
ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 713)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amgdbcm.tmc.edu
Gunnarone, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAX Plate: 17 Row: J Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6679292.
Location/Qualifiers

FEATURES

source

```

1..713
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:11430 IMAGE:3969014"
/tissue_type="Mammary tumor. Metallothionlen-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_1lib="NCI CGAP Mami"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
47..595
/codon_start=1
/product="peptidoglycan recognition protein"

```

```

BASE COUNT 172 a 218 c 178 g 145 t
ORIGIN
Query Match 43.9% Score 315; DB 10; Length 713;
Best Local Similarity 77.0%; Pred. NO. 4.4e-60;
Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```

```

QY 135 CCTGCTGAGCCCATAGTGTGCCCCGGAACGAGTGAAGGCCCTGGCATAGATGGCCCC 194
Db 90 CCTCTGAGTTTCATGCTGCCCCGAGATGATGAGGGCCCTGCCATCTCGAGTCTTA 149
QY 195 AGCAGCTAGCCTGCCCCCTTACGCTATGTGTGTGATTCGACACAGCGGGGACAGCTGCA 254
Db 150 GCCGCTGGGGGACCCAGTGTGCTACGATGATCTCACACACAGCCGACAGCTTCTGCA 209
QY 255 ACACCCCGCTCTGTCGACAGACAGAGCCCGGATGTGAGCACTACCAATGAGACAC 314
Db 210 ACAGCCGGAGCTCTCTGTGAACAGAGCCCGCATGTGACCATTAACCAAGAAATGAGC 269
QY 315 TGGGCTGTGCGAGCTGGGCTACAACTTCTGATTTGAGAGAGAGCGGCTGTATACGAG 374
Db 270 TGGGCTGTGCGAGTGTGATGATCTTCACTTCTATTTGAGAGAGAGCGGTATGTAGAG 329
QY 375 GCCGCTGTGAGTCACTTACAGGCTGCCCCATGAGTCACTTATGAAACCCATGTCATTG 434
Db 330 GCCGAGCTGTGAACATCAAGGTGTGACACAGAGCCCATCTGAGTCCATGTCTATTG 389
QY 435 GCATCAGCTTATGGGCACTACATGATGATCGGGTCCCAACCCAGGCAATCCGGGAG 494
Db 390 GCATCAGCTTATGGGCACTACATGATGATCGGGTCCCAACCCAGGCAATCCGGGAG 449
QY 495 CCCAGGCTTACTGGCCGTGTGTGCTCAGAGGAGCCCTGAGTCAACTATGTCTCA 554
Db 450 CCCAATCTTCTGGAATGTGGGTGTCTCGGGCTCTTGAATCCAAATATGAAATCA 509
QY 555 AAGCAGCCGGGATGTGACGGGTACACTCTCTCCAGGCAACCGAGCTTACCACTATCC 614
Db 510 AAGCAGCCGGGATGTGACGGGTACACTCTCTCTCCAGGCAACCGAGCTTACCACTATCC 569
QY 615 AAGATGGCCACACTACCG 633
Db 570 AAGCTGGGACACTACCG 588

```

RESULT 9

MNRNATMS1

LOCUS MNRNATMS1 678 bp mRNA linear ROD 30-JUL-1998
DEFINITION M.musculus mRNA for TAG7 protein.
ACCESSION X86374.1 GI:887524
VERSION X86374.1 GI:887524
KEYWORDS tms1 gene.
SOURCE Mus musculus.
ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 678)
AUTHORS Kustlikova, O.S., Kiselev, S.L., Borodulina, O.R., Senin, V.M., Afanas'eva, A.V., and Kabishev, A.A.
TITLE Cloning of the tag7 gene expressed in metastatic mouse tumors
JOURNAL Genetika 32 (5), 621-628 (1996)
MEDLINE 96321104
PUBMED 8755036
REFERENCE 2 (bases 1 to 678)
AUTHORS Kustlikova, O.S.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1995) O.S. Kustlikova, Inst of Gene Biology

QY 188 TGGCCCGACGACCTGAGCCTGCTTACGCTATGTTGGTGTATGACACAGCGGGGACG 247
 Db 20621 TGGCCCGACGACCTGAGCCTGCTTACGCTATGTTGGTGTATGACACAGCGGGGACG 20562
 QY 248 AGCTGCACACACCCCGCTGCTGTCAGACAGAGCCCGGAATGTGACACATACACATG 307
 Db 20561 AGCTGCACACACCCCGCTGCTGTCAGACAGAGCCCGGAATGTGACACATACACATG 20502
 QY 308 AAGACACTGGGCTGTGGAGCGTGGGCTACACTTCCTG 346
 Db 20501 AAGACACTGGGCTGTGGAGCGTGGGCTACACTTCCTG 20463
 RESULT 11
 AF154114 630 bp mRNA linear ROD 01-JUN-2000
 LOCUS Rattus norvegicus peptidoglycan recognition protein PGRP (Pgrp)
 DEFINITION Rattus norvegicus complete cds.
 ACCESSION AF154114
 VERSION AF154114.1 GI:8132325
 KEYWORDS Rattus norvegicus.
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 630)
 AUTHORS Rehman, A., Teodecki, E.E. and Krueger, J.M.
 TITLE Direct Submission
 JOURNAL Submitted (25-MAY-1999) Veterinary Comparative Anatomy,
 Pharmacology, and Physiology, Washington State University, P.O. Box
 646520, Pullman, WA 99164, USA
 FEATURES
 source
 1. 630
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /tissue_type="spleen"
 1. 630
 /gene="Pgrp"
 20. 571
 /note="similar to Homo sapiens and Mus musculus
 peptidoglycan recognition protein; similar to Trichoplusia
 n1 and Mus musculus TAG7"
 /codon_start=1
 /product="peptidoglycan recognition protein PGRP"
 /protein_id="AAF73252.1"
 /db_xref="GI:8132325"
 /translation="MLEFAMAFPPALGLADSCCFVPRSEMKALPSECKLKPVR
 VVISHYAGSCFSSPDSCQDAQNVOLYOMQLGMDVAVNFIIGEDGVYVGRGTIK
 GDHNGPIWNPMSIGITFWDGYSHRVPAKRALRALNLKCVSEGLSNVEYGHDR
 VOSTLSPDQLEYTIOSWDHYRE"
 BASE COUNT 134 a 193 c 166 g 137 t
 ORIGIN
 Query Match 41.2%; Score 295.6; DB 10; Length 630;
 Best Local Similarity 74.3%; Pred. No. 9.4e-56;
 Matches 373; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
 QY 132 CGGCTGTGAGCCCGCCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 191
 Db 63 CAGACTCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
 QY 192 CCAGACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 251
 Db 123 CCAAGGGCTGAGAAACAGTCCGCTACGTGTATCTCACACACAGCGGACAGCTTCT 182
 QY 252 GCAACACCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311
 Db 183 GCAGACGCCAGACTCTGTGAGACAGAGCGCCGCAATGTGACAGCTTACCAATGAAGAC 242

QY 312 CACTGGCTGTGGAGCTGGGCTACACACTTCTGATTGGAGAGAGCGGCTCGTATAG 371
 Db 243 AGCTGGGCTGTGGAGTGTATGCTACACACTTCTGATTGGAGAGAGTGTATGCTACG 302
 QY 372 AGGCGCGTGTGAGACTTTCACGGGTGCGCCAGCTAGTCACTTATGAACCCATGTCCA 431
 Db 303 AAGCCCGAGGCTGTGACATCAAGGGTATGACACAGAGGCGCCATGCTGAACCCCATGTCTA 362
 QY 432 TTGGCATCAGCTTATGAGGCAATACATGATGGATGGGTTGCCACACGCCAGCCATCCGGG 491
 Db 363 TGGCATCAGCTTATGAGGCAATACATGATGGATGGGTTGCCACACGCCAGCCATCCGGG 422
 QY 492 CAGCCAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 551
 Db 423 CTGCCCTAAATCTTCAATGATGGGCTGTGAGGCTTCTGAGATCTTACATATAGAG 482
 QY 552 TCAGAGACACCGGAGTGTAGAGCTACCTCTCTCCAGGACACACAGCTTACACCTCA 611
 Db 483 TCAGAGACATCGAGATGTGCAAGACTCTGTCTCAGGTGACACAGCTTACAGATCA 542
 QY 612 TCCAGAAATGGCCACACTACCG 633
 Db 543 TCCAAAGCTGGAGCCACTACCG 564
 RESULT 12
 SSC310355 259 bp mRNA linear MAM 07-DEC-2001
 LOCUS Sus scrofa partial mRNA for peptidoglycan recognition protein
 (pglyrp gene).
 ACCESSION AJ310355
 VERSION AJ310355.1 GI:17402404
 KEYWORDS peptidoglycan recognition protein; pglyrp gene.
 SOURCE Sus scrofa.
 ORGANISM Sus scrofa.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 REFERENCE 1
 AUTHORS Kappeler, S.R.
 TITLE The peptidoglycan recognition protein, PGRP, is expressed in the
 lactating mammary gland of camels
 JOURNAL Unpublished
 2 (bases 1 to 259)
 REFERENCE 2
 AUTHORS Kappeler, S.R.
 TITLE Direct Submission
 JOURNAL Submitted (06-APR-2001) Kappeler S.R., Food Science, Swiss Federal
 Institute of Technology, Zurich, LFO F26, ETHZ, CH-8092 Zurich,
 SWITZERLAND
 FEATURES
 source
 1. 259
 Location/Qualifiers
 1. 259
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /tissue_type="adipose"
 /country="Switzerland"
 2. 259
 /gene="pglyrp"
 <2. >259
 /note="pglyrp"
 /function="bacterial binding, anti-inflammatory, cytokine"
 /product="peptidoglycan recognition protein"
 /codon_start=1
 /product="peptidoglycan recognition protein"
 /protein_id="CAC83647.1"
 /db_xref="GI:17402405"
 /translation="AGSTCDNPASCROVONVQHYHRTLMCDVGNFLIGEDGLV
 EGRGMNTVGAHSGPTMNPPLSLGISFMGNMNRVPPARAIRAA"
 <2. >259
 /gene="pglyrp"
 /product="peptidoglycan recognition protein"
 BASE COUNT 47 a 86 c 83 g 43 t
 ORIGIN
 Query Match 25.8%; Score 185.4; DB 4; Length 259;
 Best Local Similarity 82.2%; Pred. No. 3.9e-31;

LOCUS	AC007085	166867 bp	DNA	linear	INV 17-MAR-2001
DEFINITION	Drosophila melanogaster, chromosome 2R, region 44D2-44E1, BAC clone BACR21H10, complete sequence.				
ACCESSION	AC007085				
VERSION	AC007085.7	GI:13374639			
KEYWORDS	HTG.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster.				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 166867) Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.Y., Buesam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreria,S., Frisze,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.				
TITLE	Sequencing of Drosophila chromosome 2R, region 44D2-44E1				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 166867) Celisner,S.E., Abdayani,A., Arcaina,T.T., Baxter,E., Blazey,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomoton,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.				
AUTHORS	Direct Submission				
TITLE	Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA				
JOURNAL	On Mar 17, 2001 this sequence version replaced gi:15670628.				
COMMENT	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (http://www.tlrl.org/sequence/) or send email to dbd@tlrl.org . Location/Qualifiers 1. 166867 /organism="Drosophila melanogaster" /strain="Y: cn bw sp" /db_xref="taxon:7227" /chromosome="2R" /map="44D2-44E1" /clone="BACR21H10 (0580)" /clone_1lb="RPCI-98 (Rosewell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"				
BASE COUNT	46731 a 36967 c 36753 g 46416 t				
QUERY	Query Match 20.5%; Score 147.4; DB 3; Length 166867; Best Local Similarity 57.3%; Pred. NO. 5.3e-23; Matches 287; Conservative 0; Mismatches 211; Indels 3; Gaps 1;				
129	ACCGGCGCTGCGAGCCCATAGTGCCTCCGACAGAGTGAAGCCCTCGCATAGAGT	188	GGCCCCAGCAGCTGAGCCTTGCCTTACGCTATGTGTGATTCGACACAGGGGGCAGCA	248	
ACATGGCCAGGCGCTGATGCTGCTCCACAGGCGAGTGGGTGCTCGCGCGCCCAAT	152119	DB 152120	GGACCTGAGCCTTGCGGCACTACTAGTATGACCATCATCCACACCGCGGCTCT	152179	
11		QY 249	GCTGCAACACCCCGCTTGTGCAGCAGCAGCCCGGATGTGACACTACACATGA	308	
11		DB 152180	ACTGCGAGACCCGTGCGCAGTCAACAGCGCGTGTGAGAGGCTCCAGAACTACACATG	152239	
11		QY 309	AGACACTGGCGTGTGCGACGCGGCTTACACTTCTGATTTGGAGAACGGGCTGTAT	368	
11		DB 152240	ACTCCCTGGCTGGCCCGCAGTGGCTATACACTTCTGATGCGGAGAGCGCAACGTGT	152299	
11		QY 369	ACGAGCGCGTGTGCTGGAATTCACGGGGTCCCACTACAGTACTATGAGACCCATGT	428	
11		DB 152300	ACGAGGGGGTGTGCTGAGAACATATGAGGGCCACAGCGC---CCAGTGGAAACCTTACA	152356	
11		QY 429	CGATTGCAATGCTTACATGAGGCACTACATGATGATGCGTGGCCACACCCAGCCATCC	488	
11		DB 152357	GCATGGCATCAGCTCTCTGCGCACTACATGAGAACACCTTGAGCGGAATGATGT	152416	
11		QY 489	CGCAGACCCAGGCTGTACTAGCGCTCGGTGCTGCTGAGAGACCCCTGAGTCCATATG	548	
11		DB 152417	CGCGCGCCAGCAGAGTGTGCTACACAGAGCGCGCTACACCTGTGCGCAGCTCGGCTACA	152476	
11		QY 549	TGCTAAGAGACACCGGATGTGACAGCTACTCTTCCAGCAACAGCTCTACACAC	608	
11		DB 152477	TGCTGTAGGCTATGCTGCGCAGGTCACCGCCACGATGCGCGGACCAATCTGGAACG	152536	
11		QY 609	TCATCCAGATGGCCACACT 629		
11		DB 152537	AGATCCGGCGCTGTGCTCCACT 152557		
11		RESULT 15			
11		AC007303/C			
11		LOCUS			
11		DEFINITION	AC007303 169509 bp DNA linear INV 21-APR-2001		
11		Drosophila melanogaster, chromosome 2R, region 44E-44F, BAC clone BACR43104, complete sequence.			
11		ACCESSION	AC007303		
11		VERSION	AC007303.7 GI:13702792		
11		KEYWORDS	HTG.		
11		SOURCE	Drosophila melanogaster.		
11		ORGANISM	Drosophila melanogaster.		
11		REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
11		AUTHORS	1 (bases 1 to 169509) Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.Y., Buesam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreria,S., Frisze,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.		
11		TITLE	Sequencing of Drosophila chromosome 2R, region 44E-44F		
11		JOURNAL	Unpublished		
11		REFERENCE	2 (bases 1 to 169509) Celisner,S.E., Abdayani,A., Arcaina,T.T., Baxter,E., Blazey,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomoton,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.		

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 12, 2002, 02:34:35 ; Search time 2689 Seconds

(without alignments)
2067.176 Million cell updates/sec

Title: US-09-462-625-4
Perfect score: 1051
Sequence: 1 MIAWALPSILRLGAAQETE.....LSPGNQYHLIONMPHYRSP 191

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09462625/runat_07112002_101525_24840/app_query.fasta.1.327
-DB=GenEmbl -GFWT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human4.0.cdi -LIST=45
-DOCLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09462625.ecgn.1.1.3637 @runat_07112002_101525_24840 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGROUVERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -WAIT -LONGLOG -DEV.TIMEOUT=120
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1051	100.0	690	9 AF076483	AF076483 Homo sapi
2	1051	100.0	724	9 AF242517	AF242517 Homo sapi
3	772	73.5	700	4 CDR131676	AJ131676 Camelus d
4	717	68.2	688	4 AY083309	AY083309 Bos tauru
5	692	65.8	549	6 AR124884	AR124884 Sequence
6	692	65.8	669	10 AF193843	AF193843 Mus muscu
7	692	65.8	680	10 AF076482	AF076482 Mus muscu
8	692	65.8	713	10 BC005582	BC005582 Mus muscu
9	667.5	63.5	630	10 AF154114	AF154114 Rattus no
10	650.5	61.9	678	10 NMNRNATMS1	X86374 M.musculus
11	505.5	48.1	166500	9 AC007785	AC007785 Homo sapi
12	441	42.0	1173	9 AY035376	AY035376 Homo sapi
13	414.5	39.4	166867	3 AC007085	AC007085 Drosophill
14	414.5	39.4	169509	3 AC007303	AC007303 Drosophill
15	414.5	39.4	172838	2 AC020451	AC020451 Drosophill
16	414.5	39.4	259336	3 AE003836	AE003836 Drosophill
17	413	39.3	1110	6 AX119918	AX119918 Sequence
18	413	39.3	1128	6 AX119915	AX119915 Sequence
19	413	39.3	1194	9 AY035377	AY035377 Homo sapi
20	405.5	38.6	652	3 AF207542	AF207542 Homo sapi
21	403.5	37.6	651	3 AY060645	AY060645 Drosophill
22	395.5	37.6	1858	9 AF242518	AF242518 Homo sapi
23	388.5	37.0	670	3 AF076481	AF076481 Trichoplus
24	386	36.7	259	4 SSC310355	AJ103555 Sus scrofa
25	379	36.1	1162	3 AF207541	AF207541 Drosophill
26	379	36.1	718	3 AF075293	AF075293 Drosophill
27	369	35.1	718	3 AF441723	AF441723 Bombyx mo
28	368	35.0	5358	4 CDR409286	AF409286 Camelus d
29	353.5	33.6	1869	3 AF500096	AF500096 Drosophill
30	348.5	33.2	1794	3 AF384856	AF384856 Homo sapi
31	348	33.1	753	6 E16480	E16480 Bombyx mori
32	347.5	33.1	590	6 AK05482	E16479 Bombyx mori
33	347.5	33.1	2751	9 AK05482	AK05482 Homo sapi
34	347	33.0	1374	3 AY058258	AY058258 Drosophill
35	347	32.9	735	3 AF313391	AF313391 Drosophill
36	346	32.9	735	3 AB016249	AB016249 Bombyx mo
37	339	32.3	622	3 AB017520	AB017520 Bombyx mo
38	339	32.3	1107	6 AX119920	AX119920 Sequence
39	339	32.3	1125	6 AX119917	AX119917 Sequence
40	338	32.2	1805	10 BC019396	BC019396 Mus muscu
41	338	32.2	159930	2 AC014935	AC014935 Drosophill
42	338	32.2	186002	3 AC006933	AC006933 Drosophill
43	338	32.2	201313	3 AC010688	AC010688 Drosophill
44	338	32.2	280887	3 AE003526	AE003526 Drosophill
45	336.5	32.0	673	3 AB017519	AB017519 Bombyx mo

RESULT 1

ALIGNMENTS

AF076483 690 bp mRNA linear PRI 15-AUG-1998
 LOCUS AF076483
 DEFINITION Homo sapiens peptidoglycan recognition protein precursor (PGRP)
 ACCESSION AF076483
 VERSION AF076483.1 GI:3342532
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 690)
 AUTHORS Kang, D., Liu, G., Lundstrom, A., Gelius, E. and Stehner, H.
 TITLE A peptidoglycan recognition protein in innate immunity conserved from insects to humans
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998) 448/8
 MEDLINE 98374308
 PUBMED 9707603
 REFERENCE 2 (bases 1 to 690)
 AUTHORS Kang, D., Liu, G., Lundstrom, A., Gelius, E. and Stehner, H.
 TITLE Direct Submission
 JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University, Stockholm S-106 91, Sweden
 FEATURES
 source location/Qualifiers
 1..690 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="bone marrow"
 1..690 /gene="PGRP"
 /gene="PGRP"
 45..635 /gene="PGRP"
 /codon_start=1
 /product="peptidoglycan recognition protein precursor"
 /protein_id="AAC31822.1"
 /db_xref="GI:3342532"
 /translation="MSRRSLMALPRLRLAQAEDTEPACCSPIVPRNEMKALAS
 ECAQHLPLRYVVSHTAGSSCNTPASCCOQARNOYHMKTLGWCVDVYNFLIGD
 GLVYEGRGNNFTGAHSGHLNPMISIGISFWGNMMDRVPQAIIRAAAGLLACGVAQGA
 LRSNVYLKGRDVRQLTSPGNQLYHLIOMPHRSP"
 45..104 /gene="PGRP"
 /note="putative"
 105..632 /gene="PGRP"
 /note="putative"
 mat.peptide /product="peptidoglycan recognition protein"
 BASE COUNT 129 a 250 c 188 g 123 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,31e-93 Length: 690
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-462-625-4 (1-191) x AF076483 (1-690)
 QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGluYlaAlaGlnGluThrGlu 20
 Db ATGCTGCTTCCCTGGGCTCTTCCAGCCCTCTCGACTCGAGCGGCTCAGAGACAGAA 119
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpIlysalAlaLeuAlaSerGlu 40
 Db GACCCGCGCTGCTGACGCCCATAGTGCCTCCGAGACAGAGGAGGCGCTGATCAGAG 179
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyrValValYalSerHisThrAlaGlySer 60
 Db TGGGCCAGACCTGAGCCCTTACGCTATGTGTGTATCGCACAGCGCGGCGAGC 239
 QY 61 SerCysAsnThrProAlaSerCysGlnGlnAlaAlaArgAsnValGlnHisTyrHisMet 80
 |||||||

Db 240 AGCTGCAACACCCCGCCCTCTGCCAGACAGAGCCCGGAATGTGCACACTACACATG 299
 QY 81 LysThrLeuGlyTrpCysAspValGlyTyrAsnPhleuLeuGlyGluAspGlyLeuVal 100
 Db AAGACACTGGCTGTGGTGGACGTGGGCTCAACTTCTCTATGGAGAAACAGGGCTCGTA 359
 QY 101 TyrGlnGlyArgGlyTyrPasnPherThrGlyAlaHisSerGlyHisLeuTrpAsnProMet 120
 Db TACGAGGCGCGGTGGTGGAACTTCAAGGCTGCCACTCAGGTCACTTATGGAACCCCATG 419
 QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 Db TCCATTGGCATCAGTTCATGAGCAACATCATGATTCGGGTCCACACCCAGCGCATC 479
 QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
 Db CGGGACGCCCGAGGGGTCTACTGCGCTGCGGTGGCTCAGAGAGCCCTGAGTCCAACTAT 539
 QY 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 Db GTGCTCAAAAGACACCGGAGATGTGACGGCTACCTCTCCAGGACACGACTTACACAC 599
 QY 181 LeuIleGlnAsnTrpProHisTyrArgSerPro 191
 Db CTCATCCAGAAATGTGGCCACTACCGCTCCCC 632
 RESULT 2
 AF242517 724 bp mRNA linear PRI 15-AUG-2000
 LOCUS AF242517
 DEFINITION Homo sapiens hypothetical protein SBB168 mRNA, complete cds.
 ACCESSION AF242517
 VERSION AF242517.1 GI:9802032
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 724)
 AUTHORS Wan, T., Zhang, W. and Cao, X.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAR-2000) Department of Immunology, Second Military Medical University & Shanghai Brilliance Biotechnology Institute, 800 Xiangyin Rd., Shanghai 200433, P.R. China
 FEATURES
 source location/Qualifiers
 1..724 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 30..620 /codon_start=1
 /product="hypothetical protein SBB168"
 /protein_id="AF99598.1"
 /db_xref="GI:9802032"
 /translation="MSRRSLMALPRLRLAQAEDTEPACCSPIVPRNEMKALAS
 ECAQHLPLRYVVSHTAGSSCNTPASCCOQARNOYHMKTLGWCVDVYNFLIGD
 GLVYEGRGNNFTGAHSGHLNPMISIGISFWGNMMDRVPQAIIRAAAGLLACGVAQGA
 LRSNVYLKGRDVRQLTSPGNQLYHLIOMPHRSP"
 BASE COUNT 161 a 251 c 186 g 126 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,44e-93 Length: 724
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-462-625-4 (1-191) x AF242517 (1-724)
 QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGluYlaAlaGlnGluThrGlu 20
 Db ATGCTGCTTCCCTGGGCTCTTCCAGCCCTCTCGACTCGAGCGGCTCAGAGACAGAA 104
 |||||||

```

Location/Qualifiers
1. 700
/organism="Camelus dromedarius"
/sub_species="Small breed"
/db_xref="taxon:9838"
/tissue_type="lactating mammary gland"
25. 606
/codon_start=1
/product="peptidoglycan recognition protein"
/protein_id="CACI9553.1"
/db_xref="GI:11990124"
/translation="MTKSCVLTWALLALISGAAREDPAGSGVTPPRKRALASEC
RERTLRPRVYVYVSHPTMGSCDPAACAQAOAVOSTYHRNLGMCDVAYNLIIDEDGIG
VYBERGNNIGAGAHGPTWAFNISIGISFTNMYNRVPPRALPAAQNLINACVVALGALNE
SNVYKCHRDVQPLPSGDGLTYHIIITWSHYA"

```

	sig_peptide	25 . 87		88 . 603	/product="peptidoglycan recognition protein"
	mat_peptide	25 . 87		678 . 683	
	POLYA_signal				
	BASE COUNT	130 a 247 c 201 g 122 t			
	ORIGIN				
Alignment Scores:					
Pred. No.:	3 51e-66				
Score:	772.00	Length:	700		
Percent Similarity:	84.74%	Matches:	140		
Best Local Similarity:	73.68%	Conservative:	21		
Query Match:	73.45%	Mismatches:	27		
DB:	4	Gaps:	2		
			1		
US-09-462-625-4 (1-191) x CDR131676 (1-700)					
QY 1 MetLeuLeuAlATPPAlaLeuProSerLeuLeuAryLeuGlyAlaAlaGlnGlu					
DB 40 GTGCTTCGTCGTGGGCGCTCTCCGCTCCGCTCCGAGCCGAGCGCGCGCA-----GAA 93					
QY 21 ASPProAlaCysCysSerProIleValProArgnsngutIrpLysAlaLeuAlaSerGlu 40					
DB 94 GACCCGGCGGCGCTCGGCTCCATCTGTCCGCCCGCCGAGAGGTGGAGGGCCCTGCTCCGAG 153					
QY 41 CysAlaGlnHisLeuSerLeuProLeuAryIleValAlaValSerHisThrAlaGlySer 60					
DB 154 TGCAGAGAAAGGGCTAACACGGCCCGCTCGCTAGCTGGTGGTGGCAGACCTCGCGGAGC 213					
QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHisMet 80					
DB 214 CACGGCAGACACCCCGCTTCTGTGGCGCCAGCAGGCCCCAGAACGTGCMAACCTACCATGTG 273					
QY 81 LysThrLeuGlyIrrPyCysAspValGlyTyrAsnPhenLeuIleGlyGluAspGlyLeuVal 100					
DB 274 CGGAACTGGGCTGTGTCGACAGTGGGCTACACTCTCTGATCGGAGAGATGGGCTCTGTG 333					
QY 101 TyrGluGlyIrrGlyIrrPasnPhenThrGlnAlaHisSerGlyHisLeuTyrPasnPromet 120					
DB 334 TAGGAAAGCCGAGGCTGGAAACATCAAGGGGGCCACGAGTCCCACTGGAAACCCATA 393					
QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetCaspArgValProThrProGlnAlaIle 140					
DB 394 TCCATATGGCATCTCTTCATAGGCAACTATATGAATCGATGCTCCCGCCCGCCGCTC 453					
QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyAlaAlaGlnGlyAlaLeuArgSerAsnTyr 160					
DB 454 CGGGCAGCCCAAGATCTGTGGCTGTGGCTGTGGCTGTGGAGCCCTGAGATCAACTAC 513					
QY 161 ValLeuLysGlyHisArgAspValGlnArgTyrLeuSerProGlyGlnGlnLeuTyrHis 180					
DB 514 GAGGTCAAGAGACACCGGGATGGAGCGCGACCCCTCTCTCCAGGATACCGGCTTACGAA 573					
QY 181 LeuIleGlnAsnThrProHisTyrArgSer 190					
DB 574 ATCATTCAGACTTGGTCACACTACCGCGGA 603					
RESULT 4					
LOCUS	AY083309	688 bp	mRNA	linear	MAY 29-MAY-2002
DEFINITION	Bos taurus oligosaccharide-binding protein mRNA, complete cds.				
ACCESSION	AY083309				
VERSION	AY083309.1	GI:19550241			
KEYWORDS					
SOURCE	Bos taurus.				
ORGANISM	Bos taurus.				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.				
REFERENCE	Yidell,C.C., Young,N., Tran,D., Yuan,J. and Selsted,M.E.				
AUTHORS	Isolation, characterization and antimicrobial properties of bovine				
TITLE					

JOURNAL	J. Biol. Chem.	277 (22)	, 19658-19664 (2002)
MEDLINE	PUBMED	22028028	11860375
REFERENCE	2 (bases 1 to 688)		
AUTHORS	Yount,N.Y., Yuan,J., Tydell,C.C. and Selsted,M.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-MAR-2002) Pathology, UC Irvine, 440D Medical Sciences		
FEATURES	Location/Qualifiers		
source	1..688		
CDS	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/cell_type="peripheral white blood cells"		
	29..601		
	/note="microbicidal; undergoes N-terminal glutamine cyclization; peptidoglycan recognition protein; PGPR"		
	/codon_start=1		
	/product="Oligosaccharide-binding protein"		
	/protein_id="FAL87002.1"		
	/db_xref="GI:19550242"		
	/translation="MSRRPPLAWLALLIGLAADDCGSIVSRGKKAALAKSCQRL RQPVVVVSHRAGSYCTPPASCCORANOVITYHREGMVDVYNLLIEDGLVYG KGNMGAHSPTNRPFLAIGISFGMTNHRVPAPSAQAASIALCGANGYLTPNYE VKGHADVQDTSPGGELTKIIIDMPHYRV"		
	order(98..100,470..470)		
	/note="disulfide bond"		
	order(146..148,281..283)		
	/note="disulfide bond"		
	order(209..211,227..229)		
	/note="disulfide bond"		
BASE COUNT	126 a 234 c 201 g 127 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	7,82e-61	Length:	688
Score:	717.00	Matches:	131
Percent Similarity:	80.75%	Conservative:	20
Best Local Similarity:	70.05%	Mismatches:	30
Query Match:	68.22%	Indels:	6
DB:	4	Gaps:	1
US-09-462-625-4 (1-191) x AY083309 (1-688)			
Oy	3 LeuaIATrPALaleProSerleuLeuaRgyLeugLyalaAlaIngLuThrGluaSppro	22	
Dd	50 CTCGCCGTGGTCTCTCTGCCCCCTCCTGGCCCTCGGGCGGCCTCAAGC-----	97	
Oy	23 AlAcysCySerProIleValPrOArysmgLuTrPyysalAlaAlaAserGluCysAla	42	
Dd	98 -----TCGGCAGCATGCTGCCCGCGGAAGTGCGGCCCTCGGCATCCAACTGACG	151	
Oy	43 GlnHisLeuSerLeuProLeuAlaGTYValValIsrHisThrAlaGlySerSerCys	62	
Dd	152 CAGAGCGTAACAACAACCCTGCTAGCTGGTGGTGTTCGCACAGCGGGCAGCGTCTGC	211	
Oy	63 AsnThrProAlaSerCysGlnGlnGlnAlaAlaArgAsnAlaGlnHisTyriHisMetLysThr	82	
Dd	212 AACACTCCGGCGCTGTGGCCAGGCAAGGCCCAAACCTGCAGTACTACCACTGCGGAG	271	
Oy	83 LeuGlyTrpCysAspValGlyTYrAsnPhelLeuIleGlyGlnSpGlyLeuValTYrGln	102	
Dd	272 CGGGCGTGTCCGACGTGGCTACAAATTCTTGATCGAGAGAATGGGCTCGTGFATGAG	331	
Oy	103 GlyArGlyTrpAsnPhetThrGlyAlaHisSerGlyHisLeuTrpAsnPrometSerIle	122	
Dd	332 GCCCGGGGCTGGAACACCTTAGGTGCTCACTCGGCCCACGTGAACCCCAATAGCCATC	391	
Oy	123 GlyIleSerPheMetGlyAsnTYrMetAspArgValPrOTHrProGlnAlaIleArgAla	142	
Dd	392 GGCAATCCCTTCATGAGCAATACATGCAATCGGGGTGCCCGGGCGCTGTCTCAGGGCG	451	
Oy	143 AlaIngInglyLeuLeuAlaCysGlyValAlaGlnGlnYalAleuArgrSerAsnTYrValLeu	162	

[illegible]

QY	181	TYTGTGCAACGAGCCGGGAGCTCTCTGTGAACAGCAGAGCCCGCAATGTGACATTCACCAAG	240
QY	81	LYSThrLeuGLIYrPCysAspValGLYTYrAsnPhcLeuIleGLYIAspGLYLeuVal	100
Db	241	AAGACTGGGCGGTGTCGATGATGACCTTCAACTTCTTATGTGGAGAGAGCGGTCAATGTC	300
QY	101	TYrGLUGLYrAspGLYTYrTrpAsnPhcThrGLYAlaHisSerGLYHisLeuTrpAsnPromet	120
Db	301	TATGAGAGCGCGGAGCGTGAACATCAACAGCGGACACAGAGGGCCCATCTCGAATCCCATG	360
QY	121	SerIleGLYIleSerPheMetGLYAsnTYrMetAspArgValProThrProGlnAlaIle	140
Db	361	TCATTGGCATCACCCTTCATGGGAACTTCATGAGACCGGGTACCCGCCAAGAGGGCCCTC	420
QY	141	ArgAlaAlaGLInGLYLeuLeuAlaCysGLYValAlaGLInGLYAlaLeuArgSerAsnTYr	160
Db	421	CGTGGCCCTTAATCTTCTGTGAAAGTGGGGGTCTCTCGGGCCCTTCTCGATTCACATCA	480
QY	161	ValLeuLeuGLYGLYHisArgAspValGLInArgThrLeuSerProGLYAsnGLInLeuTYrHis	180
Db	481	GAGTCTAAAGGACACCGGGATGTGCAAAACCTCTCTCTCCAGTGACCAACTCATATCAG	540
QY	181	LeuIleGLInAsnTrpProHisTYrArg	189
Db	541	GTCATCCAAAGCTGGGACACTACCGA	567
RESULT 7			
AF076482			
LOCUS	AF076482	680 bp	mRNA linear ROD 15-AUG-1998
DEFINITION	AF076482		Mus musculus, peptidoglycan recognition protein precursor (Pgpr)
ACCESSION	AF076482		mRNA, complete cds.
VERSION	AF076482.1	GI:3342530	
KEYWORDS			
SOURCE			Mus musculus.
ORGANISM			Mus musculus
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS			1 (bases 1 to 680)
TITLE			Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
JOURNAL			A peptidoglycan recognition protein in innate immunity conserved
MEDLINE			from insects to humans
PUBMED			Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
REFERENCE			98374308
AUTHORS			2 (bases 1 to 680)
TITLE			Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
JOURNAL			Direct Submission
MEDLINE			Submitted (08-JUL-1998) Microbiology, Stockholm University,
PUBMED			Stockholm S-106 91, Sweden
REFERENCE			1. 680
AUTHORS			Location/Qualifiers
TITLE			1. 680
JOURNAL			/organism="Mus musculus"
MEDLINE			/db_xref="taxon:10090"
PUBMED			/tissue_type="spleen"
REFERENCE			1. 680
AUTHORS			/gene="Pgpr"
TITLE			36..584
JOURNAL			/gene="Pgpr"
MEDLINE			/note="Pgpr"
PUBMED			/codon_start=1
REFERENCE			/product="peptidoglycan recognition protein precursor"
AUTHORS			/protein_id="AAC31821.1"
TITLE			/db_xref="GI:3342531"
JOURNAL			/translation="MLFACALALALGLASCSFIYPRSRMLPSECSSRLGHPPRYV
MEDLINE			VISHRTGSCFNCSPDSCQOAKNVQHTHKNELQWCDVAYNFLIGEDQHVVEGSGMNTKG
PUBMED			DTHTGFWNPMSITGTFMGFNDRVPAKRALRALNLLECQVSRGEFRSNYEYKGRHDV
REFERENCE			QSLTSGDQLYGVTSWENYRE"
AUTHORS			36..83
TITLE			/gene="Pgpr"
JOURNAL			84..581
MEDLINE			/gene="Pgpr"
PUBMED			/product="peptidoglycan recognition protein"
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS		</	

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

GAGCTGGGCTGCTGCCATGTAGCCTACACATTCCTTATTTGGAGAGGAGCGGTGATGTC 322

[illegible]

ThrLeuGlyTyrPcysAspValGlyTyrAsnPhenLeuIleGlyGluAspGlyLeuVal 100

QY 101 TyGluGlyArgGlyTyrPaspnPhetHrGlyAlaHisSerGlyHisLeuTyrPaspnProket 120
 Db 323 TATGAAGCGGAGGCTGGATCATCAAGGTTGACACACAGGCCCCATGTGAAATCCCATG 382
 QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 Db 383 TCTATTGGCATACACCTTCATCGGACATCAAGCGGATACCCGCAAGCGGCCCCCTC 442
 QY 141 ArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
 Db 443 CGTCTGCCCCCAATATCTTCTGGAATGTGGGTGTCTGGGCTTCTGTGATCCAACTAT 502
 QY 161 ValIleuGlySerGlyHisArgAspValAlaArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 Db 503 GAAGTCAAGGACACCGGATGTGCAAGACACTCTCTCTCCAGGTGACCACTATATCAG 562
 QY 181 LeuIleGlnAsnTyrProHisTyrArg 189
 Db 563 GTCATCCAAAGCTGGGAACACTACCGA 589
 RESULT 9
 AF154114 630 bp mRNA linear ROD 01-JUN-2000
 LOCUS Rattus norvegicus peptidoglycan recognition protein PGRP (Pgrp)
 DEFINITION mRNA, complete cds.
 ACCESSION AF154114
 VERSION AF154114.1 GI:8132325
 KEYWORDS
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 630)
 Rehman, A., Teedeck, E. E. and Krueger, J.M.
 Direct Submission
 Submitted (25-MAY-1999) Veterinary Comparative Anatomy,
 Pharmacology, and Physiology, Washington State University, P.O. Box
 646520, Pullman, WA 99164, USA
 Location/Qualifiers
 1..630
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /tissue_type="spleen"
 1..630
 /gene="Pgrp"
 20..571
 /gene="Pgrp"
 /note="similar to Homo sapiens and Mus musculus
 peptidoglycan recognition protein; similar to Trichoplusia
 ni and Mus musculus TAG7"
 /codon_start=1
 /product="peptidoglycan recognition protein PGRP"
 /protein_id="AAFP73252.1"
 /db_xref="GI:8132326"
 /translation="MLFAWPPALIGLADSCFYVPSSEKALPSECKGLKPYRY
 VVISHTASFCSPDSCDQARNVOLYQKQKQDVAIVNFIIGEDGHYERGMTIK
 GDHTGPIWNPMSIGITFMGDYSHRYPARAKALAAALNLKCGVSEGLRSNYEKGRHD
 VOSTLSPDQLYEIIOSMDHYRE"
 BASE COUNT 134 a 193 c 166 g 137 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.7e-56 Length: 630
 Score: 667.50 Matches: 124
 Percent Similarity: 77.37% Conservative: 23
 Best Local Similarity: 65.26% Mismatches: 34
 Query Match: 63.51% Indels: 9
 DB: 10 Gaps: 3
 US-09-462-625-4 (1-191) x AF154114 (1-630)

QY 1 MetLeuIleAlaTyrPala---LeuProSerLeuLeuArgGlyAlaAlaGlnGlyThr 19
 Db 20 ATGTGTGTTGCTGGGCTGCTCCCTGCTGCTCCCTCGGGGCGCTGGCA----- 64
 QY 20 GlnAspProAlaCysCysSerProIleValProArgAsnGlyTyrPylAlaLeuAlaSer 39
 Db 65 -----GACTCTCTGTGTTTC--GTCGTGCCCCCGACGTGAGTGAGAGCCCTCCCATGCC 115
 QY 40 GluCysAlaGlnHisLeuSerLeuProLeuArgTyrValAlaValSerHisThrIleGly 59
 Db 116 GAGTGTCTCAAGGGCGCTGAAGAAGCACTCCGCTACGGTGGATTCACACACACAGCCGCC 175
 QY 60 SerSerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisTyrHis 79
 Db 176 AGCTTCTGCACACACCCGACCTCTGTGTACAGCAGGCCCAATATGTCACACTTACCAA 235
 QY 80 MetIleThrLeuGlyTyrCysAspValGlyTyrAsnPhetHrGlyAlaHisSerGlyHisLeuTyrPaspnPro 99
 Db 236 ATGAAGACGCTGGGCTGGTGGATGTACCTTACCACTTCATGTGAGAGATGTGAT 295
 QY 100 ValTyrGluGlyArgGlyTyrPaspnPhetHrGlyAlaHisSerGlyHisLeuTyrPaspnPro 119
 Db 296 GTTACGAAAGCGGAGGCTGGATCATCAAGGTTGACACACAGGCCCCATGTGAAACCC 355
 QY 120 MetSerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAla 139
 Db 356 ATGCTTATCGGCATCACCTTCATGTGATGACTACTACACCGGATGACCTGCAAGCGGGCT 415
 QY 140 IleArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsn 159
 Db 416 CTCGCTGCTGCCCAATATCTTCTGGAATGTGGGTGTCTGGGCTTCTGTGATCCAACT 475
 QY 160 TyrValIleuGlySerGlyHisArgAspValAlaArgThrLeuSerProGlyAsnGlnLeuTyr 179
 Db 476 TATGAGGTCAAGGACATCATGATGTGCAAGACACTGTCTCTCCAGGTGACCACTTATC 535
 QY 180 HisLeuIleGlnAsnTyrProHisTyrArg 189
 Db 536 GAGATCATCCAAAGCTGGGAACACTACCGA 565
 RESULT 10
 MMRNATMS1 678 bp mRNA linear ROD 30-JUL-1998
 LOCUS M.musculus mRNA for TAG7 protein.
 DEFINITION X86374
 ACCESSION X86374.1 GI:887524
 VERSION X86374.1
 KEYWORDS tms1 gene.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.
 1 (bases 1 to 678)
 Afanas'eva, A.V. and Kabish'ev, A.A.
 Kustikova, O.S., Kiselev, S.L., Borodulina, O.R., Senin, V.M.,
 Afanas'eva, A.V. and Kabish'ev, A.A.
 Cloning of the tag7 gene expressed in metastatic mouse tumors
 Genetika 32 (5), 621-628 (1996)
 8753036
 2 (bases 1 to 678)
 Kustikova, O.S.
 Direct Submission
 Submitted (18-APR-1995) O.S. Kustikova, Inst of Gene Biology
 Russian Academy, of Science, B334, Vavilov Street 34/5, Moscow,
 RUSSIA
 3 (bases 1 to 678)
 Kiselev, S.L., Kustikova, O.S., Korobko, E.V., Prokhortchouk, E.B.,
 Kabish'ev, A.A., Iukandidin, E.M. and Georgiev, G.P.
 Molecular cloning and characterization of the mouse tag7 gene
 encoding a novel cytokine
 J. Biol. Chem. 273 (29), 18633-18639 (1998)
 98325081
 MEDLINE 9660837
 PUBMED
 COMMENT On Jul 6, 1995 this sequence version replaced gi:785010.

```

FEATURES
  source
      1..678
          location/Qualifiers
          /organism="Mus musculus"
          /strain="A/Su"
          /db_xref="taxon:10090"
          /tissue_type="adipocarcinoma"
          /dev_stage="adult"
          36..491
          /gene="tag7"
          36..491
          /gene="tag7"
          /protein_id="CA60133.1"
          /db_xref="GI:887525"
          /db_xref="MGI:107415"
          /db_xref="SPRPM:062185"
          /translation="MURFADLALGLTSCSTVPSRSEALPDESSRIGHPEVRY
          VISHATFCNSPDSCEQOARNOVHYHKNELMGCDVAYNFLIGDEGHSYRGNWINK
          DHTGPIWNPMSIGITTFMGNFMDRVKAGPCCPKSGIWSGLPEIOL"
BASE COUNT      146 a      216 c      173 g      143 t
ORIGIN
Alignment Scores:
  Pred. No.:      2,31e-54      Length:      678
  Score:      650.50      Matches:      125
  Percent Similarity:      77.25%      Conservative:      21
  Best Local Similarity:      66.14%      Mismatches:      35
  Query Match:      61.89%      Indels:      9
  DB:      10      Gaps:      1
US-09-462-625-4 (1-191) x MMRNATMS1 (1-678)
QY      1 MetLeuAlaATrPALaProSerLeuLeuArgLeuGlyAlaAlaGlnGlnLurGln 20
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      36 ATGTGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 83
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      21 AspProAlaCysCysSerProIleValProArgGlnGlnLurPlyAlaLeuAlaSerGlu 40
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      84 -----TTCAGTTTCATGTCGCCCGCCGAGTGAAGGAGGCGCTGCATCCGAG 131
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      41 CysAlaGlnHisLeuSerLeuProLeuArgGlyValValAlaSerHisThrAlaGlySer 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      132 TGTCTGAGCCCGCTGGGGACCCAGTTCGTCAGTGTGATCTGCACACAGCCCGCAGC 191
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      61 SerCysAnthrProAlaSerCysGlnGlnGlnAlaArgAnValGlnHisThrHisMet 80
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      192 TTCTGCAACAGCCCGGAGCTCTGTGAACAGCAGCCCGCAATGTGCACATTACCAAG 251
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      81 LysThrLeuGlyTrpCysAspValGlyTrpAsnPhenLeuIleGlyGluAspGlyLeuVal 100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      252 AATGAGCTGGGCTGTGCGATGTGAGCTTCAACTTCTTATTGGAGAGCGGTGATGTC 311
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      101 TyrGlnGlyArgGlyTrpAsnPhenThrGlyAlaHisSerGlyHisLeuTrpAsnProMet 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      312 TATGAAGCGCGAGCTGGAACATCAAGGTCACACAGAGGCCCATCTGGAATCCCATG 371
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      372 TCTTTTGGCAGACCTTCATGGGAACCTTCATGAGACCGGGTAC--GCAAGCGGGCCCTC 429
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      430 CGTCTGCTCCCTAAATCTCTTCGATCTGGGGTGTCTCGGGCTTCCTGAGATCCAACTAT 489
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      161 ValLeuGlyGlyHisAspArgValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      490 GAGTCAAGAGCAGCGGAGATGTGCAAGACCTCTCTCCAGTGACCACTATATAG 549
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      181 LeuIleGlnAsnTrpProHisTyrArg 189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      550 GTCTATCAAGAGCTGGGAACATACGA 576
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 11

```

```

AC007785/c
LOCUS      AC007785      166500 bp      DNA      linear      perl 11-JUN-1999
DEFINITION Homo sapiens chromosome 19, BAC 282485 (CIT-B-344H19), complete
            sequence.
ACCESSION      AC007785
VERSION      AC007785.1      GI:5042403
KEYWORDS
SOURCE
ORGANISM      Homo sapiens.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS      Lander, J.E., McCready, P.M., Richardson, P., Sakai, D.,
            (bases 1 to 166500)
            Lander, J.E., McCready, P.M., Richardson, P., Sakai, D.,
            Butkhart-Schultz, K., Gordon, L., Scott, D., Johnson, G., Stillwagen, S.,
            Phil, H., Velasco, N., Do, L., Regala, W., Terry, A., Dangnan, L.,
            Eiler, A., Christensen, M., Georgescu, A., Avila, D., Altix, C.,
            Andreise, T., Amico-Keller, G., Coefield, D., Duarte, S., Lucas, S.,
            Bruce, R., Thomas, P., Quan, G., Krommiller, B., Atellano, A.,
            Sanders, C., Ow, D., Nolan, M., Trong, S., Kobayashi, A., Olsen, A.S. and
            Carrano, A.V.
            Sequence analysis of a 2.5 Mb region in 19q13.2-13.3 between APOE
            and CSRI
            Unpublished
            2 (bases 1 to 166500)
            Lander, J.E.
            Direct Submission
            Submitted (11-JUN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
            Map and sequence oriented from q centromere to telomere. BAC 282485
            overlaps cosmid R30477 to the left from bases 1 to 4,419 of this
            accession, and overlaps BAC 264576 (CIT-B-297N14; AC006262) from
            bases 146,203 to 166,500. Additional chromosome 19 map and sequence
            information may be obtained at:
            http://www.bio.lnl.gov/dbir/genome/genome.html.
            Location/Qualifiers
            1..166500
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="19"
            /map="19q13.3 between DM and CSRI"
            /clone="CIT-B-344H19"
            /clone_11b="Cal Tech CIT-B BAC library"
            /note="LNL clone name: BC282485"
            70..145
            /rpt_family="(TGAA)n"
            complement(177..453)
            /rpt_family="Aluub"
            complement(472..712)
            /rpt_family="MIR"
            join(<746..904,1157..1239,4294..4451,5875..6036,
            12784..12994,13080..13162,13803..13958,14053..14128,
            14429..14568,14726..14804,14982..15039,15813..15885,
            16025..16122)
            /note="Hypothetical arginine-rich gene product;
            Hypothetical CDS constructed from overlapping ESTs and
            x-ray predictions. EST coverage is lacking for some
            portions of the CDS. Gene may be alternatively spliced, as
            multiple transcripts map into this region; alternatively,
            these individual clones may represent artifactual events
            during RT and cDNA creation"
            /codon_start=3
            /evidence=not_experimental
            /product="BC282485.1"
            /protein_id="AAB38244.1"
            /db_xref="GI:5042403"
            /translation="TLAMDOPAGLOVDYFVGVHAYRVNVSQVLELEVEDRMTADQ
            WKGEFDAGFIDLTHTKGNFKOFNIFCHMLESALTOSSESYTLDTLTVDLESNRK
            MGRPSGLAPRSALNSKRYLILYSVEIDHIVLPYOGKPPVYLOGITRSIKE
            ELGRLOGIDGONTRENETIWHLRDOYSRLASRRLEADOLGSRERALAGRAROE
            AEALRGVRLLELROBRGIGHRVAGRGDDCRRLAKELBEAKASESLKRLKLTLT
            SELALYRGGKRSRGKRPAPSPSPGGRALRDPAFVFAKERKOREILKQOORN
            RLGGSGSDSPSVMSROTQPPALVTGDAIPNRSRNSSSVDSFRSSASCSIDL
            EDFESISLRGCHRRRGRKPPSPWGSNMKSPVPRSHHOKSLANSGMWPIKEYSSE

```

```

misc_feature
      HOADMATIDARLALQRYMNRDMS"
      746..904
      /note="DSD similarity to overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (12..170): 99% identity.-AA283147 z149d07.r1 NCI_CGAP_GCB1
      Homo sapiens cDNA clone IMAGE:713101 5'; (6..163); 99%
      identity.-AA143160 z149d07.r1 Soares pregnant uterus NBHPU
      Homo sapiens cDNA clone 505261 5'; (1..43); 100%
      identity.-AA142881 z149d07.s1 Soares pregnant uterus NBHPU
      Homo sapiens cDNA clone 505261 3'; (570..527); 64%
      identity."
      1157..1239
      /note="Predicted exon, program: graal2exons_human_1.3,
      frame: 0, quality: good, score: 62.000-DSD similarity to
      overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (171..255): 95% identity.-AA283147 z149d07.r1
      NCI_CGAP_GCB1 Homo sapiens cDNA
      clone IMAGE:713101 5'; (164..247); 92% identity.-AA143160
      z149d07.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA
      clone 505261 5'; (44..126); 100% identity.-N28707
      yx66d11.r1 Homo sapiens cDNA clone 266709 5'; (1..66); 98%
      identity.-AA142881 z149d07.s1 Soares pregnant uterus NBHPU
      Homo sapiens cDNA clone 505261 3'; (526..438); 82%
      identity."
      1536..1634
      /rpt_family="(CA)n"
      complement(1673..1973)
      /rpt_family="AluSx"
      repeat_region
      complement(2040..2185)
      /rpt_family="LINE2"
      repeat_region
      complement(2200..2498)
      /rpt_family="AluSx"
      repeat_region
      complement(2576..2746)
      /rpt_family="AluSg/x"
      repeat_region
      complement(2747..3044)
      /rpt_family="AluSx"
      repeat_region
      3048..3750
      /rpt_family="MER64"
      3841..4060
      /rpt_family="MER64"
      4294..4451
      /rpt_family="MER64"
      /note="DSD similarity to overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (256..438); 81% identity.-AA283147 z149d07.r1
      NCI_CGAP_GCB1 Homo sapiens cDNA
      clone IMAGE:713101 5'; (248..400); 99% identity.-AA143160
      z149d07.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA
      clone 505261 5'; (127..281); 97% identity.-N28707
      yx66d11.r1 Homo sapiens cDNA clone 266709 5'; (67..222);
      98% identity.-AA142881 z149d07.s1 Soares pregnant uterus
      NBHPU Homo sapiens cDNA clone 505261 3'; (437..275); 97%
      identity."
      complement(4789..5083)
      /rpt_family="AluSx"
      5875..5972
      /note="DSD similarity to overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (439..506); 71% identity.-AA143160 z149d07.r1 Soares
      pregnant uterus NBHPU Homo sapiens cDNA clone 505261 5';
      (282..379); 100% identity.-AA142881 z149d07.s1 Soares
      pregnant uterus NBHPU Homo sapiens cDNA clone 505261 3';
      (274..175); 95% identity."
      5875..5898
      /note="DSD similarity to N28707 yx66d11.r1 Homo sapiens
      cDNA clone 266709 5'; (223..246); 100% identity."
      5906..6036
      /note="Predicted exon, program: graal2exons_human_1.3,
      frame: 1, quality: excellent, score: 100.000"
      complement(6148..6443)
      /rpt_family="AluSx"
      repeat_region
      complement(6608..6648)
      /rpt_family="LINE2"

repeat_region
      complement(7118..7340)
      /rpt_family="MIR"
      repeat_region
      complement(7538..7639)
      /rpt_family="(CA)n"
      repeat_region
      complement(7792..8054)
      /rpt_family="MIR"
      repeat_region
      complement(8789..9083)
      /rpt_family="AluJo"
      repeat_region
      9456..9520
      /rpt_family="MER5B"
      complement(9539..9841)
      /rpt_family="AluSx"
      repeat_region
      9909..9944
      /rpt_family="AluSx"
      repeat_region
      10099..10401
      /rpt_family="MER5B"
      repeat_region
      /rpt_family="AluSx"
      complement(10587..10752)
      /rpt_family="AluJ"
      repeat_region
      complement(10753..11057)
      /rpt_family="AluSx"
      repeat_region
      complement(11065..11191)
      /rpt_family="AluJ"
      repeat_region
      11322..11619
      /rpt_family="AluSg"
      repeat_region
      11775..11814
      /rpt_family="Ar_r1ch"
      repeat_region
      complement(11864..11885)
      /rpt_family="Ar_r1ch"
      repeat_region
      complement(11916..11976)
      /rpt_family="AluS"
      repeat_region
      complement(12055..12088)
      /rpt_family="Ar_r1ch"
      repeat_region
      complement(12186..12327)
      /rpt_family="LINE1"
      repeat_region
      complement(12334..12384)
      /rpt_family="LINE2"
      12784..12994
      /rpt_family="LINE2"
      /note="Predicted exon, program: graal2exons_human_1.3,
      frame: 1, quality: excellent, score: 92.000"
      12847..12908
      /rpt_family="GC_r1ch"
      12895..12994
      /rpt_family="GC_r1ch"
      /note="DSD similarity to AA765166 n279c05.s1 NCI_CGAP_GCB1
      Homo sapiens cDNA clone IMAGE:1301672; (222..123); 97%
      identity."
      13002..13032
      /rpt_family="GC_r1ch"
      13080..13162
      /note="Predicted exon, program: graal2exons_human_1.3,
      frame: 2, quality: excellent, score: 100.000-DSD
      similarity to AA765166 n279c05.s1 NCI_CGAP_GCB1 Homo
      sapiens cDNA clone IMAGE:1301672; (122..48); 100%
      identity."
      complement(13799..13954)
      /rpt_family="GC_r1ch"

repeat_region

Alignment Scores:
Pred. No.: 1,45e-37 Length: 166500
Score: 505.50 Matches: 100
Percent Similarity: 73.72% Conservative: 1
Best Local Similarity: 72.99% Mismatches: 0
Query Match: 48.10% Indels: 36
DB: 9 Gaps: 1

US-09-462-625-4 (1-191) x AC007785 (1-166500)
QY 91 AsnPhenLeuIleGlyGluAspGlyLeuValTyrtGluYArgGlyTrpAsnPhenThrGly 110
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 17384 AGCTTCTGTATTGGAGAAAGCGGCTGCTATPACAGAGGCGGTGCTGGAACCTTACCGGT 17325

QY 111 AlAhISerGlyHisLeuTrpAsnPrometSerIleGlyIleSerPhenMetGlyAsnTrp 130
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 17324 GCCCAGTCAGAGTCATTATGGAACCCCATCTCCATGTGGCAATCAGTCACTGAGCAATC 17265

```


[illegible]

Wed Nov 13 09:37:09 2002

us-09-462-625-4.rge

Page 13

OY 188 TYRArgSer 190
: : : : : :
DB 150729 TGGAGGCC 150737

Search completed: November 12, 2002, 05:02:37
Job time : 2802 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 00:51:04 : Search time 85 Seconds

(without alignments)
93.200 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1 MLAMALPSLRGAAQETE.....LSPGNQLYHLQNPYRSP 191

Sequence: 1051

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1051	100.0	196	1	PGRP_HUMAN
2	692	65.8	182	1	PGRP_MOUSE
3	388.5	37.0	182	1	PGRP_TRINI
4	345.5	32.9	196	1	PGRP_BOMMO
5	127	12.1	150	1	NAAA_BPT7
6	127	12.1	150	1	NAAA_BPT7
7	86.5	8.2	481	1	ATZB_PSED
8	79.5	7.6	267	1	SURE_HELPJ
9	79.5	7.6	267	1	SURE_HELPJ
10	77.5	7.4	356	1	YGTJ_ECOLI
11	75.5	7.2	443	1	FBIA_CRIGR
12	75	7.1	687	1	VS41_GIALA
13	75	7.1	764	1	GLGB_STRAU
14	75	7.1	803	1	MDDC_CHLTC
15	74	7.0	713	1	TS44_GIALA
16	74	7.0	1477	1	HTK7_HYDAT
17	73.5	7.0	439	1	YSCN_YEREN
18	73.5	7.0	439	1	YSCN_YEREN
19	73.5	7.0	1733	1	VNUA_PRYCA
20	72.5	6.9	378	1	CARA_PSEAE
21	72	6.9	207	1	SNAT_SHEEP
22	72	6.9	408	1	G58B_DROME
23	72	6.9	1286	1	IRBP_BOVIN
24	71.5	6.8	362	1	IRBP_BOVIN
25	71.5	6.8	444	1	IRBP_BOVIN
26	71	6.8	402	1	GLGA_DEIRA
27	71	6.8	775	1	LIPS_HUMAN
28	71	6.8	1120	1	YB85_YEAST
29	71	6.8	1123	1	DC13_DROME
30	71	6.8	1589	1	DC13_DROME
31	70.5	6.7	443	1	FBIA_MOUSE
32	70.5	6.7	610	1	GIDA_CHLTR
33	70.5	6.7	3312	1	CLR3_HUMAN

ALIGNMENTS

RESULT 1	PGRP_HUMAN	STANDARD:	PRT:	196 AA.
ID	PGRP_HUMAN			
AC	075594			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Peptidoglycan recognition protein precursor (SBB168).			
GN	PGLYRP OR PGRP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.			
RC	TISSUE=bone marrow;			
RA	MEDLINE=98374508; PubMed=9707603;			
RA	Kang D., Liu G., Lundstroem A., Gellius E., Steiner H.;			
RT	"A peptidoglycan recognition protein in innate immunity conserved from insects to humans.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998). <i>Ang 18</i>			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Lamerdin J.E., McCready P.M., Richardson P., Sakalasis G.;			
RA	Burthart-Schultz K., Gordon L., Scott D., Johnson G., Stillwagen S.;			
RA	Phan H., Velasco N., Do L., Regala W., Terry A., Dangnan L.;			
RA	Erlar A., Christensen M., Georgescu A., Avila J., Attix C.;			
RA	Andreise T., Amico-Keller G., Coefield J., Duarte S., Lucas S.;			
RA	Brue R., Thomas P., Quan G., Krommiller B., Ariellano A., Sanders C.;			
RA	OW D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Man T., Zhang W., Gao X.;			
RL	Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN INNATE IMMUNITY.			
CC	- FUNCTION: THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO (BY SIMILARITY).			
CC	- SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-ASSOCIATED FORMS (BY SIMILARITY).			
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BONE MARROW, WEAK EXPRESSION FOUND IN KIDNEY, LIVER, SMALL INTESTINE, SPLEEN, THYMUS, PERIPHERAL LEUKOCYTE, LUNG AND FETAL SPLEEN.			
CC	- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AF076483; AAC31822.1; -			

DR EMBL: AC007785; AAD38243.1; -
 DR EMBL: AF242517; AAF95958.1; -
 DR HSSP: P00806; IABO.
 DR Genew: HGNC:8904; PGLYRP.
 DR MIM: 604963; -
 KW Immune response; Cytokine; Apoptosis; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 196 PEPTIDOGLYCAN RECOGNITION PROTEIN.
 FT DISULFID 67 73 POTENTIAL.
 FT CARBOHYD 112 112 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 196 AA; 21731 MW; D954C51440DC27DC CRC64;
 Query Match 100.0%; Score 1051; DB 1; Length 196;
 Best Local Similarity 100.0%; Pred. No. 5; 3e-94;
 Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLAMALPSILRLGAAQETEDPACCSPIVPRNEMKALASECAQHLSPRLRYVVSHTAGS 60
 DB 6 MLAMALPSILRLGAAQETEDPACCSPIVPRNEMKALASECAQHLSPRLRYVVSHTAGS 65
 QY 61 SCNTPASCCOQARNVQYHMKTLGMCVDVYNFLIGEDGLYEGRGMTGASHGLNMPM 120
 DB 66 SCNTPASCCOQARNVQYHMKTLGMCVDVYNFLIGEDGLYEGRGMTGASHGLNMPM 125
 QY 121 SIGISFMGNMTDRVPTPOAIRAAQGLIACGVAQGLRSNYLKGHRDVRTLSFGNQLYH 180
 DB 126 SIGISFMGNMTDRVPTPOAIRAAQGLIACGVAQGLRSNYLKGHRDVRTLSFGNQLYH 185
 QY 181 LIONMPTHRSP 191
 DB 186 LIONMPTHRSP 196
 RESULT 2
 PGRP MOUSE STANDARD; PRT; 182 AA.
 ID PGRP MOUSE STANDARD; PRT; 182 AA.
 AC 088393; 062185;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Peptidoglycan recognition protein precursor (Cytokine tag7).
 GN PGLYRP OR PGRP OR TAG7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_Taxid-10090;
 OX [1]
 RN RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE-Spleen;
 RX MEDLINE-98374308; PubMed-9707603;
 RX Kang D., Liu G., Lundstrom A., Gellus E., Steiner H.;
 RT "A peptidoglycan recognition protein in innate immunity conserved from
 insects to humans";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).
 RN [2]
 RN RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE-98325081; PubMed-9660837;
 RX Kiselev S.L., Kustikova O.S., Korobko E.V., Prokhorchouk E.B.,
 RA Kabishhev A.A., Lukandin E.M., Georgiev G.P.;
 RT "Molecular cloning and characterization of the mouse tag7 gene
 encoding a novel cytokine";
 RL J. Biol. Chem. 273:18633-18639(1998). *July 17*
 RN [3]
 RN RP SEQUENCE FROM N.A.
 RA Slayton W.B., Rigas A., Hancock J.D., Zaugg J.K., Le T.V.,
 RA Trantman M.S., Spangrude G.J., Carroll W.L., Schibler K.R.;
 RT "Granulocyte-colony stimulating factor up-regulates expression of
 murine tag7 during myeloid differentiation";
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J; TISSUE-Small intestine;
 RX MEDLINE-21085660; PubMed-11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Atakawa T., Hara A., Fukunishi Y., Konno H., Aichi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kanaya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombers P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Welt C., Whitaker C., Wilmink L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RN Nature 409:685-690(2001).
 RN [5]
 RN RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN
 CC INNATE IMMUNITY. THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO.
 CC -1- SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-
 CC ASSOCIATED FORMS.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN AND LUNG. ALSO
 CC DETECTED IN BRAIN AND THYMUS. IN THE LUNG, EXPRESSED IN THE
 CC INTRAALVEOLAR SPACE. IN THE BRAIN, EXPRESSED IN THE PURKINJE CELLS
 CC OF THE CEREBELLUM AND IN CERTAIN LAYERS OF NEURONS IN THE
 CC HIPPOCAMPUS. ALSO DETECTED IN CELLS FILLING THE SPACE WITHIN THE
 CC INTESTINAL VILLOS.
 CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
 CC FAMILY.
 CC [6]
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL: AF076482; AAC31821.1; -
 DR EMBL: AF193843; AAF06335.1; -
 DR EMBL: AK008335; BAB25611.1; -
 DR EMBL: BC005582; AAH05582.1; -
 DR EMBL: X86374; CAA60133.1; ALT_SEQ.
 DR EMBL: Y12088; CAA72803.1; -
 DR MGD; MGI:1345092; Pglyrp.
 FT IMMUNE response; Cytokine; Apoptosis; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 182 PEPTIDOGLYCAN RECOGNITION PROTEIN.
 FT DISULFID 54 60 POTENTIAL.
 SQ SEQUENCE 182 AA; 20489 MW; 9844E2137F047F14 CRC64;
 Query Match 65.8%; Score 692; DB 1; Length 182;
 Best Local Similarity 67.2%; Pred. No. 1; 6e-59;
 Matches 127; Conservative 21; Mismatches 33; Indels 8; Gaps 1;
 QY 1 MLAMALPSILRLGAAQETEDPACCSPIVPRNEMKALASECAQHLSPRLRYVVSHTAGS 60
 DB 1 MLAMALPSILRLGAAQETEDPACCSPIVPRNEMKALASECAQHLSPRLRYVVSHTAGS 65
 QY 61 SCNTPASCCOQARNVQYHMKTLGMCVDVYNFLIGEDGLYEGRGMTGASHGLNMPM 120
 DB 66 SCNTPASCCOQARNVQYHMKTLGMCVDVYNFLIGEDGLYEGRGMTGASHGLNMPM 125
 QY 121 SIGISFMGNMTDRVPTPOAIRAAQGLIACGVAQGLRSNYLKGHRDVRTLSFGNQLYH 180
 DB 126 SIGISFMGNMTDRVPTPOAIRAAQGLIACGVAQGLRSNYLKGHRDVRTLSFGNQLYH 185

OY 181 LIONMPHYR 189
Db 173 VIOSMEHYR 181

RESULT 3

PGRP_TRINI STANDARD; PRT: 182 AA.

AC 075537; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptidoglycan recognition protein precursor.

OS Trichoplusia ni (Cabbage looper).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Noctuidae; Noctuidae; Plusiinae; Trichoplusia.
NCBI_Taxid=7111;

[1]

RP SEQUENCE FROM N.A., SEQUENCE OF 17-21, TISSUE SPECIFICITY, AND

RP INDUCTION.

RC TISSUE-SPECIFICITY: PUBMED-9707603;

RA Kang D., Liu G., Lundstroem A., Gellius E., Steiner H.;

RT "A peptidoglycan recognition protein in innate immunity conserved from

RT insects to humans.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).

CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND TRIGGERS THE

CC PROPOLOXIDASE CASCADE WHICH IS AN IMPORTANT INSECT INNATE IMMUNE

CC DEFENSE MECHANISM.

CC -1- SUBUNIT: MONOMER (PROBABLE).

CC EXPRESSION OBSERVED IN HEMOCYTE. NO EXPRESSION DETECTED IN

CC GOT.

CC -1- INDUCTION: BY BACTERIAL CHALLENGE.

CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

DR EMBL; AF076481; AAC31820.1; -

RESULT 4
PGRP_BOMMO STANDARD; PRT: 196 AA.

AC 09XTND; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptidoglycan recognition protein precursor.

OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Bombycoidea; Bombycidae; Bombyx.
NCBI_Taxid=7091;

[1]
RP SEQUENCE FROM N.A., SEQUENCE OF 24-72; 99-118; 147-159 AND 183-196,

RP TISSUE SPECIFICITY, AND INDUCTION.

RC STRAIN-Kinshu x Showa; TISSUE-Pat body;

RA MEDLINE-99223509; PubMed-10207004;

RT "A pattern recognition protein for peptidoglycan. Cloning the cDNA and

RT the gene of the silkworm, Bombyx mori.";

RL J. Biol. Chem. 274:11854-11858(1999).

CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND TRIGGERS THE

CC PROPOLOXIDASE CASCADE WHICH IS AN IMPORTANT INSECT DEFENSE

CC MECHANISM.

CC -1- SUBUNIT: MONOMER (PROBABLE).

CC EXPRESSION OBSERVED IN FAT BODY,

CC EPITHELIAL CELLS AND HEMOCYTES. NOT DETECTED IN MALPIGHIAN

CC TUBULES, SILK GLAND OR MIDGUT.

CC -1- INDUCTION: BY BACTERIAL CHALLENGE.

CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

DR EMBL; AB016605; BAA77210.1; -

OS	Pseudomonas sp. (strain ADP).
OG	Plasmid pADP-1.
OC	Bacteria; Proteobacteria.
OX	NCBI_TaxID=4760;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=92208208; PubMed=9055410;
RA	Boudry-Mills K.L., de Souza M.L., Mandelbaum R.T., Wackett L.P.,
RD	Sadowatz M.J.:
RT	"The atzB gene of Pseudomonas sp. strain ADP encodes the second
RL	enzyme of a novel atrazine degradation pathway."
CC	Appl. Environ. Microbiol. 63:916-923(1997)."
CC	-1- FUNCTION: TRANSFORMS HYDROXYATRAZINE TO N-ISOPROPYLAMMELIDE
CC	(DIHYDROXY-ISOPROPYL-ATRAZINE).
CC	-1- CATALYTIC ACTIVITY: 4-(ethylanino)-2-hydroxy-6-(isopropylamino)-
CC	1,3,5-triazine + H ₂ O = N-isopropylammelide + ethylamine.
CC	-1- PATHWAY: Atrazine degradation; second step.
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC	-1- SIMILARITY: BELONGS TO THE ATZ/TRZ FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/announce/
CC	or send an email to license@lsb.slb.ch).
DR	EMBL: U066917: AAC45138.1; -
DR	InterPro: IPR002604; ATZ_TRZ.
DR	Pfam: PF01685; ATZ_TRZ; 1.
KW	Hydrolase; Plasmid.
SQ	SEQUENCE 481 AA; 52113 MW; BF2EF1C5891CD4A1 CRC64;
<hr/>	
Query Match 8.2%; Score 86.5; DB 1; Length 481;	
Best Local Similarity 23.6%; Pred. No. 0.72; Matches 62; Indels 25; Gaps 5;	
Matches 33; Conservative 20; Mismatches 62; Indels 25; Gaps 5;	
QY	26 SPIVPRNEWMKALASECAQHLSLPURVVVSHTAGSCNTPRACCOOARNOHNYHNKTIGW 85
Db	219 SPVIAYETPEFVESARLARLRHLGSL-----HTHGGEETPRM-----VARFERSLDW 265
QY	86 CDGVYNFLIGDGLVYEGRCNFTGAASGHLMNMISIGISMGYMDRVPPQAIRAAQC 145
Db	266 CE--NRGFVGVD--VWLHAGEFTPAADIARL-----RATGTGANCPAPVPLVGAEV 313
QY	146 LLACGVAQCALRSNVYLKGH 165
Db	314 TDIPTMAAAGVRVGFVDGH 333
<hr/>	
RESULT 8	
SURE_HELPJ	
ID SURE_HELPJ STANDARD; PRT; 267 AA.	
AC O9ZKSQ;	
DT 30-MAY-2000 (Rel. 39, Created)	
DT 30-MAY-2000 (Rel. 39, Last sequence update)	
DE 15-JUN-2002 (Rel. 41, Last annotation update)	
DE Acid phosphatase sure (EC 3.1.3.2).	
SURE GN JHP0865.	
OS Helicobacter pylori j99 (Campylobacter pylori j99)	
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;	
OC Helicobacter	
OX NCBI_TaxID=85963;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=99120557; PubMed=9923682;	
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,	
RA Smith D.R., Noonan B., Guild B.C., deGeorge B.L., Carmel G.,	
RA Tummino P.J., Caruso A., Gira-Nickelsen M., Mills D.M., Ives C.,	
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyts G.F.,	
RT Trust T.J.:	
TT "Genomic sequence comparison of two unrelated isolates of the human	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: U18997; AAA57880.1; -
 DR EMBL: AE000390; AAC7611.1; -
 DR Ecogene: EG12721; y9jJ.
 DR Hypothetical protein; signal; complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 356 HYPOHETICAL PROTEIN YGJ.
 SQ SEQUENCE 356 AA; 40099 MW; 51D6089E9284385 CRC64;

Query Match 7.4%; Score 77.5; DB 1; Length 356;
 Best Local Similarity 21.3%; Pred. No. 3.7;
 Matches 46; Conservative 17; Mismatches 76; Indels 77; Gaps 11;

OY 35 KALASECAQHLSTPLRYVYVSHTAGSSC-----NTPASCCQAARVQHYHMKTLGMC 86
 DB 2 KLITPCRLALPFCYAF--SAGEEARPAHDDTKTPTITSTSPSFRFYGLGVG-- 57
 OY 87 DVGYNFLIGED-----GLYEGRGWNTGASHG-----HLWN 118
 DB 58 --GYWDLSEGENHKKYSDGYIEGLEMKYGSWFGILY--GEGWTYQVQDHGNMVPVPSHG 114
 OY 119 PMSIIS-FMGNYMDRVPPPAIRAAQGLACGVAQALRS-----NYLAKGR 166
 DB 115 GFEIGINRGYGYRTNDGTETMLSRQ-----DSSLDLQWMDFTPDGLGVIPNTR 166
 OY 167 DVQRFLS-----PGNQLYHLI-----QNMPPH 188
 DB 167 DMTALKYQNLSGNFRYVYATPAGHDESKAMLEP 202

RESULT 11
 FBL4-CRIGR
 ID FBL4-CRIGR STANDARD: PRT: 443 AA.
 AC 055058;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE EGF-containing fibulin-like extracellular matrix protein 2 precursor
 DE (Fibulin-4) (FBL-4) (H411 protein).
 GN EFEMP2 OR FBLN4.
 OS Citicellus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Citicellus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RA Heine H., Delude R.L., Monks B., Golenbock D.T.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: AF046870; AAC03101.1; -
 DR HSSP: P00736; IAPQ.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.

DR Pfam: PF00008; EGF: 4.
 DR SMART: SM00179; EGF_CA: 4.
 DR SMART: SM00001; EGF_Like: 2.
 DR PROSITE: PS00010; ASX_HYDROXYL: 4.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 4.
 DR PROSITE: PS01187; EGF_CA: 6.
 DR Repeat: EGF-like domain; Calcium-binding; Glycoprotein; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 443 EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR
 FT DOMAIN 36 81 MATRIX PROTEIN 2.
 FT DOMAIN 123 163 EGF-LIKE 1, DIVERGENT.
 FT DOMAIN 164 202 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 203 242 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 243 282 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 283 328 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 127 140 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 134 149 BY SIMILARITY.
 FT DISULFID 151 162 BY SIMILARITY.
 FT DISULFID 168 177 BY SIMILARITY.
 FT DISULFID 173 186 BY SIMILARITY.
 FT DISULFID 188 201 BY SIMILARITY.
 FT DISULFID 207 217 BY SIMILARITY.
 FT DISULFID 213 226 BY SIMILARITY.
 FT DISULFID 228 241 BY SIMILARITY.
 FT DISULFID 247 258 BY SIMILARITY.
 FT DISULFID 254 267 BY SIMILARITY.
 FT DISULFID 269 281 BY SIMILARITY.
 FT DISULFID 287 300 BY SIMILARITY.
 FT DISULFID 294 309 BY SIMILARITY.
 FT DISULFID 315 327 BY SIMILARITY.
 FT CARBOHYD 198 198 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 198 198 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 443 AA; 49432 MW; 0BCFE5D7323D95E5 CRC64;

Query Match 7.2%; Score 75.5; DB 1; Length 443;
 Best Local Similarity 30.0%; Pred. No. 7.4;
 Matches 24; Conservative 11; Mismatches 26; Indels 19; Gaps 4;

OY 2 LLMALPSLRLGAA--QETEDPACCSPIVPRNEMKALASECAQHLSTPLRYVYVSHTAG 59
 DB 12 LLMAL-LLLLLGASBPDESDSTYECDDYEMDASQHCND-----V 55
 OY 60 SSCNT-PASCCQAARVQHY 78
 DB 56 NECLTFPEACKGEMKCNHY 75

RESULT 12
 VS41-GIALA
 ID VS41-GIALA STANDARD: PRT: 687 AA.
 AC P92127;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Variant-specific surface protein VSP41 precursor (CRISP-90).
 OS Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
 OX NCBI_TaxID=5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=02-4A1.
 RC MEDLINE=97321554; PubMed=9178264;
 RA Papanastasiou P., Brudefer T., Li Y., Bommeil C., Koehler P.;
 RA Primary structure and biochemical properties of a variant-specific
 RA surface protein of Giardia.";
 RL Mol. Biochem. Parasitol. 86:13-27(1997).
 RN [2]
 RP CHARACTERIZATION.
 RP MEDLINE=97333006; PubMed=9078242;
 RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
 RA "The variant-specific surface protein of Giardia, VSP41, is a

RT glycosylated and palmitoylated protein.";
RL Biochem. J. 322:49-56(1997).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -1- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
CC AT THE REDUCING TERMINUS.
CC -1- PTM: PALMITOYLATED.
CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z83743; CAB06038.1; -.
DR HSSP: O14763; ID06.
DR GlycoSuiteDB: P92127; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR005127; Giardia_VSP.
DR Pfam: PF03302; VSP; 2.
DR SMART: SM00181; EGF; 2.
DR SMART: SM00001; EGF-like; 1.
DR SMART: SM00261; Fu; 3.
DR Antigen: Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
KW Signal.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 687 VARIANT-SPECIFIC SURFACE PROTEIN VSPA41.
FT DOMAIN 15 660 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 661 681 POTENTIAL.
FT DOMAIN 682 687 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Query Match 7.1%; Score 75; DB 1; Length 687;
Best Local Similarity 22.4%; Pred. No. 14;
Matches 46; Conservative 19; Mismatches 72; Indels 68; Gaps 12;

QY 14 GAQEDPDA-----C-CPPIVPRNEMKALASBGAHLSPLRVVYVSHTGSSCNTPAS 67
DB 403 GCABCTREPGLKCTCKP---NRKPAGTSD-----NTTCREKT---CEDTIV 445
QY 68 CQO-----QARNVQHYHMKTLGMDVGYNFLIGEDGLVEGRGWNFTGAHSHLM 117
DB 446 CGGTSGACDAIVIDANGKEHY---CSYCGETNKPFI--DGLCTDNKGTN--ACCTDHTC 498
QY 118 NPASIG-ISMGN--YDRP-----TPQAIRAAGLIAC 149
DB 499 SYCAAGFELWGGCYKIDTVPGSYMCKSTYAGVCDTPNANRRFVVPVKAISQSYLAC 558
QY 150 GVAQALRSNMYLKGHDVORTLSP 174
DB 559 GNPGLTLAGNNAVGVGCSQCTAP 583

RESULT 13
GLGB STRAU STANDARD; PRT; 764 AA.
AC P52960;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (Glycogen branching
DE enzyme).
GN GLGB.
OS Streptomyces aureofaciens.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1894;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 10762 / CCM 3239;
RX MEDLINE-94347823; PubMed-8068720;
RA Honerova D., Kotmanec J.;
RT "Cloning of the putative glycogen branching enzyme gene, glgB, from
RT Streptomyces aureofaciens.";
RL Biochim. Biophys. Acta 1200:334-336(1994).
CC -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
CC glycogen.
CC -1- PATHWAY: Glycogen biosynthesis; third step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L11647; AAA67437.1; -.
DR InterPro: IPR000461; Alpha_amylase.
DR InterPro: IPR004193; Isoamylase_N.
DR Pfam: PF00128; alpha-amylase; 1.
DR Pfam: PF02922; isoamylase_N; 1.
KW Glycogen biosynthesis; Transferase; Glycosyltransferase.
FT ACT_SITE 440 440 BY SIMILARITY.
FT ACT_SITE 493 493 BY SIMILARITY.
FT ACT_SITE 561 561 BY SIMILARITY.
SQ SEQUENCE 764 AA; 85325 MW; 6B45482E4A268ACF CRC64;

Query Match 7.1%; Score 75; DB 1; Length 764;
Best Local Similarity 32.6%; Pred. No. 15;
Matches 29; Conservative 8; Mismatches 30; Indels 22; Gaps 6;

QY 90 YNFL-IGEDLVYEGRS-----NMFPGA---HSG-----HLNPMISIGISFMG--NY 130
DB 137 YRFLPALGELDLHLIGSRHHELMTALGSPHEHGAVGTRETVPANALGVRYGDPFV 196
QY 131 MDRVPTPQAIRAAGC---LLACGVAQAL 156
DB 197 WDVAAYPMRSLGASGWMELFLPGVAEGAL 225

RESULT 14
MDDC_CHLTR STANDARD; PRT; 803 AA.
ID MDDC_CHLTR
AC O84767;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE MURC/dcl bifunctional enzyme [includes: UDP-N-acetylmuramate--alanine
DE ligase (EC 6.3.2.8) (UDP-N-acetylmuramoyl-L-alanine synthetase); D-
DE alanine-D-alanine ligase (EC 6.3.2.4) (D-alanylalanine synthetase)]
DE (D-Ala-D-Ala ligase)].
GN MURC/DDL OR CT762.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/DW-3/Cx;
RX MEDLINE-9900809; PubMed-9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1- FUNCTION: CELL WALL FORMATION.
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl + L-alanine -> ADP +
CC phosphate + UDP-N-acetylmuramoyl-L-alanine.

```

CC -1- CATALYTIC ACTIVITY: ATP + D-alanine + D-alanine -> ADP + phosphate
CC + D-alanyl-D-alanine.
CC -1- PATHWAY: Peptidoglycan biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE MURCEP
CC FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE D-ALANINE-
CC D-ALANINE LIGASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001348; AAC68357.1; -.
CC HSP: P07862; I10M.
CC InterPro: IPR000291; Data_Data_Ligase.
CC InterPro: IPR000713; Mur_Ligase.
CC InterPro: IPR004101; Mur_Ligase_C.
CC Pfam: PF01225; Mur_Ligase; 1.
CC Pfam: PF01820; Data_Data_Ligase; 1.
CC Pfam: PF02875; Mur_Ligase_C; 1.
CC TIGR: TIGR01082; murC; 1.
CC TIGR: TIGR01205; Data_Data_Ligase; 1.
CC PROSITE: PS00843; DATA_DATA_LIGASE_1; 1.
CC PROSITE: PS00844; DATA_DATA_LIGASE_2; FALSE_NEG.
CC Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
CC ATP-binding; Multifunctional enzyme; Complete proteome.
CC FT DOMAIN 1 446 UDP-N-ACETYLTRANSFERASE--ALANINE LIGASE.
CC FT NP_BIND 447 803 ATP (POTENTIAL).
CC FT NP_BIND 111 117 ATP (POTENTIAL).
CC SQ SEQUENCE 803 AA; 89230 MW; 82B8523BF1F4C1C8 CRC64;

Query Match 7.1%; Score 75; DB 1; Length 803;
Best Local Similarity 25.0%; Pred. No. 16;
Matches 29; Conservative 24; Mismatches 41; Indels 22; Gaps 7;

QY 86 CDVGNFLIGEDG-----LVYEGRWNN-FTGAHSHLMNPMISGIFMGNYMDRVPPT 137
DB 220 CLOGHTF--GIDSSCDLHILSYQEGWRLYFTAKYQDVYADIEVLGMHVNLA--- 274

QY 138 QATRAAGL-LACVAGCALNSVYKGRDVOQLSPGN--QLXHLIOMPHYYS 190
DB 275 ----AAGIALSLGIDEGAIRNAF--RGFSVQRRLORKNSSEPLFLEDYAHPS 324

RESULT 15
TS4_GIALA STANDARD; PRT; 713 AA.
ID TS4_GIALA
AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major surface-labeled trophozoite antigen 417 precursor.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 30957 / WB;
RX MEDLINE=90280395; PubMed=2352929;
RA Gillin F.D., Hagblom P., Harwood J., Alez S.B., Reiner D.S.,
RA McCaffery M., So M., Guiney D.G.;
RA "Isolation and expression of the gene for a major surface protein of
RT Giardia lamblia.",
RL Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
RN [2]
RP SEQUENCE OF 480-620 FROM N.A.
RC STRAIN-AD-1;

```

```

RX MEDLINE=93314970; PubMed=8325510;
RA Ey P.L., Mayrhofer G.;
RT "Two genes encoding homologous 70-kDa surface proteins are present
RT within individual trophozoites of the binucleate protozoan parasite
RT Giardia intestinalis."
RL Gene 129:257-262(1993).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -1- DOMAIN: CONTAINS 29 REPEATS OF THE CXC MOTIF.
CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M33641; AAA02688.1; -.
CC EMBL: M97488; AAA02581.1; -.
CC PIR: A35502; A35502.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR005127; Giardia_VSP.
CC Pfam: PF03302; VSP; 2.
CC SMART: SM00181; EGF; 1.
CC SMART: SM00001; EGF-like; 1.
CC SMART: SM00261; FU; 3.
CC Signal: Antigen; Glycoprotein; Transmembrane; Repeat.
CC FT CHAIN 1 17
CC FT SIGNAL 1 17
CC FT DOMAIN 18 713 MAJOR SURFACE-LABELLED TROPHOZOITE
CC FT TRANSMEM 680 708 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 709 713 CYTOPLASMIC (POTENTIAL).
CC FT CARBOHYD 289 289 N-LINKED (GLCNAc... ) (POTENTIAL).
CC FT CARBOHYD 676 676 N-LINKED (GLCNAc... ) (POTENTIAL).
CC FT VARIANT 582 582 A -> T (IN STRAIN ADELAIDE-1).
CC FT VARIANT 606 606 A -> S (IN STRAIN ADELAIDE-1).
CC SQ SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;

Query Match 7.0%; Score 74; DB 1; Length 713;
Best Local Similarity 22.6%; Pred. No. 18;
Matches 45; Conservative 19; Mismatches 85; Indels 50; Gaps 9;

QY 2 LLAVALPSLLRLGAQETEDPAC--CSPYVPRNEMKALASECAQHLSLPLRYVVVSHTAG 59
DB 6 LLAIVYLQIARFACIQEADDGKCKTGCTIGDTW---CSECGANAYAPVNGGCVDNAE 62

QY 60 SSCNTPASCCQARNVQHYHMKTLGMC-DVGVNPLIGEDGLVYEGRWNTGAHSHLMN 118
DB 63 GPSKT--LCPOH-----SAGKTCQCGNSFWKDCVSSGREL-----PGH--- 101

QY 119 PMSIGSFNGNWDNRVPTP-----QAIRAAGCLLCGVAGCALNSVYKGRDVOQL 172
DB 102 --SLCLSSDGDGCTAAAGCFAPVAGANTQGVSIACGGDTGV-----TI 144

QY 173 SPGNQLYHLIOMPHYRSP 191
DB 145 AAGGNYYKGIADCAECSAP 163

```

Search completed: November 12, 2002, 02:23:54
 Job time : 87 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 01:41:04 : Search time 144 Seconds
(without alignments)
273.299 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051
Sequence: 1 MLTAMALPSILRLGAAQETE.....LSPGNQLYHLIQQNPHYRSP 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPPREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	772	73.5	193	6	Q9GK12
2	717	68.2	190	6	Q8SP7
3	667.5	63.5	183	11	Q9JLN4
4	441	42.0	341	4	Q96L89
5	414.5	39.4	184	5	Q9V4X2
6	413	39.3	373	4	Q96L88
7	405.3	38.6	185	5	Q9V3B7
8	403.5	38.4	185	5	Q95S09
9	386	36.7	86	6	Q8WME9
10	379	36.1	203	5	Q9VYX7
11	371.5	35.3	368	4	Q9HD75
12	369	35.1	195	5	Q8HSZ1
13	353.5	33.6	500	5	Q8T5Q2
14	348.5	33.2	576	4	Q96PD5
15	347.5	33.1	634	4	Q96N74
16	347	33.0	345	5	Q9VXN9

Result No.	Score	Query Match	Length	ID	Description
17	339	32.3	194	5	Q9BL11
18	336.5	32.0	208	5	Q9BL12
19	334	31.8	195	5	Q97369
20	330	31.4	186	5	Q9VS97
21	329.5	31.4	190	5	Q9V997
22	329.5	31.4	530	11	Q8VCS0
23	328	31.2	529	11	Q9QX22
24	327.5	31.2	369	5	Q8SXQ7
25	327	31.1	337	5	Q8VSW0
26	317	30.2	215	5	Q9VGN3
27	315	30.0	500	11	Q9QX21
28	289.5	27.5	182	5	Q9VY96
29	248.5	23.6	611	5	Q9VSV9
30	229.5	21.8	280	5	Q9VSV8
31	229.5	21.8	299	5	Q9GK7
32	224	21.3	539	16	Q86334
33	221	21.0	520	5	Q9GK5
34	181.5	17.3	138	5	Q9GK6
35	177.5	16.9	416	16	Q9KZ5
36	151	14.4	222	16	Q8XK3
37	148.5	14.1	308	16	Q9S2P9
38	147.5	14.0	282	5	Q9GK97
39	147.5	14.0	505	5	Q9V9P7
40	133.5	12.7	304	16	Q8XLA4
41	126	12.0	151	9	Q9T132
42	91	8.7	879	10	Q9FZ8
43	85.5	8.1	642	13	P79941
44	82	7.8	268	5	O62353
45	80.5	7.7	281	11	O88812

ALIGNMENTS

RESULT 1

Q9GK12 PRELIMINARY: PRT: 193 AA.

AC Q9GK12: 01-MAR-2001 (TREMBLrel. 16, Created).

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Peptidoglycan recognition protein precursor.

GN PCRP.

OS Camelus dromedarius (Dromedary) (Arabian camel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.

OX NCBI_TaxID=9838;

RN [1]

RP SEQUENCE FROM N.A.

KC TISSUE=LACTATING MAMMARY GLAND;

RA Kappeler S.R., Farah Z., Puhon Z.;

RT "Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan Recognition Protein.";

RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Kappeler S.R., Farah Z., Puhon Z.;

RT mammary gland of camels and binds to lactic acid bacteria.";

RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AJ131676; CAC10553.1; -;

DR EMBL: AJ409286; CAC84130.1; -;

KW Signal.

FT SIGNAL. 1 21

FT CHAIN. 22 193

SO SEQUENCE 193 AA; 21377 MW; B6A1BD818030A7CB CRC64;

PEPTIDOGLYCAN RECOGNITION PROTEIN.

Query Match Score 772; DB 6; Length 193;

Best Local Similarity 73.7%; Pred. No. 1.le-70;

Matches 140; Conservative 21; Mismatches 27; Indels 2; Gaps 1;

1 MLTAMALPSILRLGAAQETDPACCSPIYPRNWKALASCAQHLSTPLRYVYVSHRAGS 60

```

Db      6 VLLVALLALALSLGAAR--EDDPAGCSIVPRREWRMLASECHERLRLTRPVRYVVSHTAGS 63
QY      61 SCNTFASCSQQQARNVQHTHMKTLGMCVGYNFLIGEDGLVYGRGWNFTGASGHLNPM 120
Db      64 HCDPTFASCSQQQANQVQSHVNLGMCVGYNFLIGEDGLVYGRGWNFTGASGHPWTNPI 123
QY      121 SIGISFMGNYMDRVPPTPOAIRAAGGLACGVAGALRSNYVLKGRHDVORTLSPGNOLPH 180
Db      124 SIGISFMGNYMNRVPPRRLRAAQNLLACGVAGALRSNYEVKGRHDVPTLSPDRLYE 183
QY      181 LIQNPHTRS 190
Db      184 IIQTMSTHYRA 193

RESULT 2
O8SP7
ID      O8SP7      PRELIMINARY:      PRT:      190 AA.
AC      O8SP7
DT      01-JUN-2002 (TREMBLrel. 21, Created)
DT      01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
OC      Oligosaccharide-binding protein.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Tytell C.C., Yount N.Y., Tran D., Yuan J., Seisted M.E.;
RT      Isolation, characterization, and antimicrobial properties of bovine
RT      oligosaccharide-binding protein: A microbicidal granule protein of
RT      eosinophils and neutrophils."
RT      J. Biol. Chem. 0:0-0(2002).
DR      EMBL: AY083309; AAL87002.1; -
SQ      SEQUENCE 190 AA: 21063 MW; 2BA7D659438F4ED7 CRC64;

Query Match      68.2%; Score 717; DB 6; Length 190;
Best Local Similarity 70.1%; Pred. No. 4.2e-65;
Matches 131; Conservative 20; Mismatches 30; Indels 6; Gaps 1,

QY      3 LAMALPILLRLGAQETDEDPACCSPIVPRNEKALASECAOHLSLPLRVVVSHTAGSSC 62
Db      8 LAMVLLALGLGLAADP-----CGSIVSRGKMGALASKCSQRLRPVRYVVSHTAGSYC 61
QY      63 NTPASCSQQQARNVQHTHMKTLGMCVGYNFLIGEDGLVYGRGWNFTGASGHLNPMPSI 122
Db      62 NTPASCSQQQANQVQSHVNLGMCVGYNFLIGEDGLVYGRGWNFTGASGHPWTNPI 121
QY      123 SIGISFMGNYMDRVPPTPOAIRAAGGLACGVAGALRSNYVLKGRHDVORTLSPGNOLPH 182
Db      122 SIGISFMGNYMNRVPPRRLRAAQNLLACGVAGALRSNYEVKGRHDVPTLSPDRLYE 181
QY      183 QNMPHYR 189
Db      182 QQMPHYR 188

RESULT 3
O9JLN4
ID      O9JLN4      PRELIMINARY:      PRT:      183 AA.
AC      O9JLN4
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE      01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE      Peptidoglycan recognition protein PGRP.
GN      PGRP.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]

```

RP	SEQUENCE FROM N.A.					
RC	STRAIN-SPRAGUE-DAWLEY;	TISSUE=SPLEEN;				
RA	Reiman A., Teodecki E.E., Krueger J.M.;					
RH	Submitted (MAY-1999)	to the EMBL/Genebank/DBJ databases.				
DR	AF154114; AAU73252.1;	-				
SQ	SEQUENCE 183 AA; 20590 MW; 5B9C1B7AA8A2EC21 CAC64;					
Query Match		63.5%; Score 667.5; DB 11:	Length 183;			
Best Local Similarity		65.3%; Pred. No. 4.4e-60;				
Matches 124; Conservative 23; Mismatches 34; Indels 9; Gaps 3,						
OY	1 MILANA-LPGLRLCAOETEDPDACCSPVYPNEKKALASGCAQLSLPLRYVVVSHTAG 59					
	: :	:	: : : : : : : : : : : : : : : :			
DB	1 MLFAAPAPPALLGLTA-----DSCCF--VPPESEWKALPSECSSKGLKRPRVRYVISHTAG 52					
OY	60 SSCNTFPASCQQQAARVQYHYHMTLTLCMDGYNYFLIGEGLYVEGRGMNFTGAHSCHLNMP 119					
	: : : : : : : : :					
DB	53 SFCCSPPDCCEQDAARVOLYOXMKOLCMCDVANYNFLIGDEGHVEYEGRWITKGDTPIPNMP 112					
OY	120 MSIGTSPFGNVNDRYDPFPOAIRAAGLLACGVAGALSNNYLTKGRHDVQPTLSGNOLDY 179					
	: : :					
DB	113 MSGTGTEMGDYSHRPVKAKRALPALNLKCGVSEGFLSNSNEYKGRDVQSFTLSPGDOLY 172					
OY	180 RLIGNMPHYR 189					
	: :					
DB	173 ELIQSWDHTR 182					

ID	096LBB	PRELIMINARY:	PRT:	341 AA.
AC	096LBB			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Peptidoglycan recognition protein-I-alpha precursor			
GN	PGIYRPIALPHA			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21433985; PubMed=11461926;			
RA	Liu C, Xu Z, Gupta D, Dziarski R.			
RT	"Peptidoglycan recognition proteins, a novel family of four human			
RT	innate immunity pattern recognition molecules."			
RL	J. Biol. Chem. 276:34686-34694(2001).			
DR	EMBL: AY035376; AAK72484.1; ..			
KW	Signal.			
FT	SIGNAL			
EQ	SEQUENCE 341 AA; 37611 MW; 8ADD5AA97B632076 CRC64;			
		POTENTIAL		
		1 17		
	Query Match	42.0%; Score 441; DB 4; Length 341;		
	Best Local Similarity	41.4%; Pred. No. 1.1e-36;		
	Matches 77; Conservative 43; Mismatches 60; Indels 6; Gaps 3;			
QY	8	BSLLALGAQAEETEDA----CCSPYPRREKAKALSECQHSLSPLRYVYVSHTAGSSCN	63	
DB	157	PLLLKEETCLDQHHPMPKRYCPNIIKRSAMVARETHCPK-MNLPAKYVIIITHAGTSCT	215	
QY	64	TPASCOOARVQNHMKTLGWCDDVGYNLFLEDDLVYEGRCQMNFTGASHGILAMPMSIG	123	
DB	216	VSTDOGVYVRNIGSFHMDIRNCFDGIHFLVQDQDGGVYEGVGMHIGQSITF-IGFNDAIG	274	
QY	124	ISFMGNVMDRNPVTPAIRAAGCLLAGCAVAGALRSNIVYLKGRDVOFTLSPGNQLYHLIQ	183	
DB	275	IAFIYFEVKEPNNAAALEAADLIQCAVEVGYLTENYILMGSHSDVNVNIIISPCQALYNIIS	334	
QY	184	NMPHYR 189		
DB	335	TWPHKR 340		

```

RESULT 5
09V4X2 PRELIMINARY; PRT; 184 AA.
AC 09V4X2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE CG14745 protein.
GN GGRP-SC2 OR CG14745.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Beres P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kessison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclio J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrstkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003836; AAC59051.1; -.
DR HSSP; P00806; 1LBA.
DR FlyBase; FBgn0043575; GGRP-SC2.
SQ SEQUENCE 184 AA; 19829 MM; 0F99D04914B07238 CRC64;

Query Match 39.4%; Score 414.5; DB 5; Length 184;
Best Local Similarity 47.2%; Pred. No. 2.5e-34;
Matches 77; Conservative 31; Mismatches 54; Indels 1; Gaps 1;

```

```

OY 148 ACYAGALRSNYVLKGRDVOFTLSPGNOLYHLIONMPHYRS 190
DB 142 SDVSRGQIVSGYLLIGHROVGSTECPGTINMEIFTSWMA 184

RESULT 6
096LB8 PRELIMINARY; PRT; 373 AA.
AC 096LB8;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Peptidoglycan recognition protein-I-beta precursor.
GN PGLTP1BETA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21433985; PubMed=11461926;
RA Liu C., Xu Z., Gupta D., Dzidarski R.;
RT "Peptidoglycan recognition proteins, a novel family of four human
RT innate immunity pattern recognition molecules."
RL J. Biol. Chem. 276:34686-34694(2001).
DR EMBL; AY035377; AAK72485.1; -.
DR InterPro; IPR02086; Aldehyde_dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN.1.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 373 AA; 40662 MM; 148BA166018A66AA CRC64;

Query Match 39.3%; Score 413; DB 4; Length 373;
Best Local Similarity 40.0%; Pred. No. 8.8e-34;
Matches 76; Conservative 42; Mismatches 66; Indels 6; Gaps 3;

```

```

OY 4 ANAPSLR-----LGAQETEDPACCSPIYPRNEMKALASECAHSLPRLVYVSHAG 59
DB 185 SYVQPLVKGENCELAPROKTSLKACPGVPRSRVWCAREFHCBP-MTLPAKYGIIHTAG 243
OY 60 SSCNTPASCOOQARNOYHMKTLGMCDDVGVNPLIGEDGLVYGRGNFNGASHGLMP 119
DB 244 RHCNIDECRLVRODQSTYIDPKSCDIGNLVGDGAIYEGVGNVGS-STPRTD 302
OY 120 MSIGISFMGNMVRVETPQAIRAAGLACGVAAGALRSNYVLKGRDVOFTLSPGNOLY 179
DB 303 IALGIFMGFTGCTPPNAALAEAOOLICAMVKGITLPYLLVGHSDVARTLSPGOLY 362
OY 180 HLIQNPYHR 189
DB 363 NIISTWPHRK 372

RESULT 7
09V3B7 PRELIMINARY; PRT; 185 AA.
AC 09V3B7;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE CG8577 protein and CG14746 protein (Peptidoglycan-recognition protein-
DE SC1B).
GN GGRP-SC1A OR CG8577.3 OR CG14746.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;

```

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Peiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abiri J.F., Abdayani A., An H.-J., Andrews-Plannkoc H., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchman M.R., Bouck J., Brokstein P., Bottler P.,
 RA Burts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mitsuhashi N.V., Mobarry C., Morris J., Mostreil A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Switzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT The genome sequence of *Drosophila melanogaster*.;
 RN Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20558582; PubMed-11106397;
 RA Werner T., Liu G., Kang D., Ekengren S., Steiner H., Hultmark D.,
 RT "A family of peptidoglycan recognition proteins in the fruit fly
 RT *Drosophila melanogaster*."

DB Proc. Natl. Acad. Sci. U.S.A. 97:13772-13777(2000).
 DR EMBL: AE003836; AAF59052.1; -;
 DR EMBL: AE003836; AAF59054.1; -;
 DR EMBL: AF207543; AAG23736.1; -;
 DR FlyBase: FBgn0033327; PGRP-SC1b.
 DR FlyBase: FBgn0043576; PGRP-SC1a.
 SQ SEQUENCE 185 AA; 20395 MW; F23F8D80A33541AC CRC64;

Query Match 38.6%; Score 405.5; DB 5; Length 185;
 Best Local Similarity 49.1%; Pred. No. 2,1e-33;
 Matches 79; Conservative 20; Mismatches 61; Indels 1; Gaps 1;
 QY 28 IYPRNEMKALASECAOHLSPRLRYVYVSHTAGSSCNPASCOOQARNOVHYHMKTLGMD 87
 DB 24 VYSAEMGGRGAKWTGVLGNYLAIHHTAGSYCETRAACNAVLOSVQYHNDLSLMPD 83
 QY 88 VGVNPLIGEDGLVEGRGNTGASHGLNPMSTGISFGKNTMDRVTPQAIRAAGGL 147
 DB 84 IGVNPLIGDGNVEGRKWNMGAAAE-WNPYSIGISFGLNMYDLEPNMISAQOOL 142
 QY 148 AGVAGALRSNVYLKGRDQVORTLSPGNOLYHLIQNMPHY 188
 DB 143 NDVNRGOLSSGITLYGHROYSAETPCGTHIMETIRGMSHW 183

RESULT 8
 QY 095S09 PRELIMINARY; PRT; 185 AA.
 AC 093S09;

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE GH07464P.
 GN PGRP-SC1b OR CG8577.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Phryganea; Neoptera; Emdopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RA Stapleton M., Brokstein P., Hong L., Abdayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Paclet J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY060645; AAT28193.1; -;
 DR FlyBase: FBgn0033327; PGRP-SC1b.
 SQ SEQUENCE 185 AA; 20336 MW; A83F8D80A33541AF CRC64;

Query Match 38.4%; Score 403.5; DB 5; Length 185;
 Best Local Similarity 49.1%; Pred. No. 3,4e-33;
 Matches 79; Conservative 19; Mismatches 62; Indels 1; Gaps 1;
 QY 28 IYPRNEMKALASECAOHLSPRLRYVYVSHTAGSSCNPASCOOQARNOVHYHMKTLGMD 87
 DB 24 VYSAEMGGRGAKWTGVLGNYLAIHHTAGSYCETRAACNAVLOSVQYHNDLSLMPD 83
 QY 88 VGVNPLIGEDGLVEGRGNTGASHGLNPMSTGISFGKNTMDRVTPQAIRAAGGL 147
 DB 84 IGVNPLIGDGNVEGRKWNMGAAAE-WNPYSIGISFGLNMYDLEPNMISAQOOL 142
 QY 148 AGVAGALRSNVYLKGRDQVORTLSPGNOLYHLIQNMPHY 188
 DB 143 NDVNRGOLSSGITLYGHROYSAETPCGTHIMETIRGMSHW 183

RESULT 9
 Q8WME9
 ID Q8WME9 PRELIMINARY; PRT; 86 AA.
 AC Q8WME9;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Peptidoglycan recognition protein (Fragment).
 GN PGLYRP.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 NX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADIPOSE TISSUE;
 RA Kappeler S.R.;
 RT "The peptidoglycan recognition protein, PGRP, is expressed in the
 RT lactating mammary gland of camels."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ10355; CAC83647.1; -;
 FT NON_TER 1
 FT CHAIN 1
 FT NON_TER 1
 FT 86
 FT 86
 SQ SEQUENCE 86 AA; 9517 MW; BFC5DC02AE11CBES CRC64;

Query Match 36.7%; Score 386; DB 6; Length 86;
 Best Local Similarity 76.7%; Pred. No. 7,8e-33;
 Matches 66; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
 QY 58 AGSSCNTPASCOOQARNOVHYHMKTLGMDVGVNPLIGEDGLVEGRGNTGASHGLN 117
 DB 1 AGSTCDNPASCRQOVQVHYHVTILHMCVDGVNPLIGEDGLVEGRGNTGASHGLN 60

QY 118 NPMISGISEMGNYMDRYPPOAIRAA 143
 DB 61 NPLSLGISFMONTMRRVPPAIRRAA 86

RESULT 10
 09VYX7 PRELIMINARY; PRT: 203 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CG11709 protein (Peptidoglycan-recognition protein-Sa)
 DE (A130827D)
 GN pGRP-SA OR CG11709.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; Pubmed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gockayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brinkstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Eyringelista C.C., Ferrara C., Ferrelle S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwan C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen R.,
 RA Shue B.C., Siden-Kiamos I., Simpson M.C., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter R., Venter A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissendbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BP CL CN BW;
 RX MEDLINE=20558582; Pubmed=11106397;
 RA Werner T., Liu G., Kengren S., Steiner H., Hultmark D.,
 RT "A family of peptidoglycan recognition proteins in the fruit fly
 Drosophila melanogaster."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:13772-13777(2000).
 RN [3]
 RP SEQUENCE FROM N.A.

RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R.A., Gonzalez M., Guerin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungai C.J., Nuccio J., Paclet J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (Jan-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AF003486; AAF48056.1; -
 DR EMBL, AF207541; AAG23735.1; -
 DR EMBL, AF207540; AAG23734.1; -
 DR EMBL, AY075293; AAL68160.1; -
 DR FLYBase; FBgn030310; pGRP-SA.
 DE SEQUENCE 203 AA; 22260 MW; D200A6EA79C66731 CRC64;

Query Match 36.1%; Score 379; DB 5; Length 203;
 Best Local Similarity 40.1%; Pred. No. 1.2e-30;
 Matches 76; Conservative 34; Mismatches 76; Indels 2; Gaps 2;

QY 3 LAMALPSLLRLGAAGETEDPACCSPVPRNEMKALASECAQHSLRLRYVYVSHTGSSC 62
 DB LVLLALFAVSACKSRQ-RSPANCPITKLRQWGGKPSLGLHVOVRPRYVVIHHTVGGC 74

QY 63 NTPASCOQCARVVOHYHMKTLGCDVGYNFLIGEDLVYEGRGNFTGAHSGHIANPMSI 122
 DB 75 SELLCAEELIOMQAVHOMELDFNDISYFLIGNDSIVYEGTGWLRGANT-YGYNAIGT 133

QY 123 GISFMGNTMDRYPPOAIRAAQGLACGYAAGALNSNYLKGHRDVOFTLSFGNQLYHLI 182
 DB 134 GIAFGNFEVDKLPSPALQAARDLACGVQGESLSDYALIGSQYISQSPCLTYLNET 193

QY 183 QNMPHYRS 190
 DB 194 QEMPHYLS 201

RESULT 11
 09HD75 PRELIMINARY; PRT: 368 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Hypothetical 40.0 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Zhang W., Wan T., Cao X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AF242518; AAF95959.1; -
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 368 AA; 40020 MW; 1E7497073A5DAFD CRC64;

Query Match 35.3%; Score 371.5; DB 4; Length 368;
 Best Local Similarity 44.1%; Pred. No. 1.5e-29;
 Matches 64; Conservative 33; Mismatches 47; Indels 1; Gaps 1;

QY 45 LSLPRVYVSHFASGSCNTPASCOQCARVVOHYHMKTLGCDVGYNFLIGEDLVYEGR 104
 DB 224 MTLPAKYGIIHTAGRTCTISDECRILVDISFYIDRLKSCDIGVFLVGCGAIYEGV 283

QY 105 GGNFGAHSCHLAMPMSISIFMGNTMDRYPPOAIRAAQGLACGYAAGALNSNYLVC 164
 DB 284 GNNVGS-STPEYDIALAGITMGFTTGTPNAALAEAAQDLQCAVMAGYLTPTLLVG 342

QY 165 HRDVOFTLSFGNQLYHLIOMPHYR 189
 DB 343 HSDVARTLSPGOALYNIISTWPHK 367

RESULT 12

O8WSZ1

ID O8WSZ1 PRELIMINARY; PRT; 195 AA.

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)

DE Peptidoglycan recognition protein.

OS Bombyx mori (Silk moth).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;

OC Bombycoidea; Bombycidae; Bombyx.

OX NCBI_Taxid=7091;

RN [1]

RP SEQUENCE FROM N.A.

RA Kim S.H., Lee H.S., Kim J.W., Lee Y.S., Ryu K.S., Jun Y., Minoru Y.,

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF441723; AAL32058.1; -

SQ SEQUENCE 195 AA; 21478 MW; 17F85F01BDA035E CRC64;

Query Match

Best Local Similarity 35.1%; Score 369; DB 5; Length 195;

Matches 76; Conservative 35; Mismatches 66; Indels 20; Gaps 4;

OY 1 MLTAMALPSL-----RLGAOETEDPACCSPVPRNEMKALASECAQHLSPLRY 51

DB 1 MLVA---PSLLLVFLVSTETLNAASEDGE-----IPTEWSTESRRKQPLKSPIDL 50

OY 52 VVVSHTAGSSCNTPASCOOQARNOHYHMKTLGMCVDYGNFLIGEDGLYEGRGMTFGA 111

DB 51 VVIOHTVSNDCFTDEBCLLSVNSLRHMRHLAGPKDGLGSFVAGNGKIKIEGAGMNNIGA 110

OY 112 HSGHLMNPKSIGSPMGNTYDRVTPQAIAGAGLACGAGALSNVYLKGRDYORT 171

DB 111 HTLH-YNNISIGIGFGDFEKLPTQALQAVDFLACGVENMLTBDYHYVGHQQLINT 169

OY 172 LSPGNOLYHLIQNMPHY 188

DB 170 LSPGATVLSIESPMHY 186

RESULT 13

O8T5O2

ID O8T5O2 PRELIMINARY; PRT; 500 AA.

DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)

DE Peptidoglycan recognition protein-1C isoform x.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Choe K.-M., Werner T., Stoven S., Hultmark D., Anderson K.V.,

RT Requirement for a Peptidoglycan Recognition Protein (PGRP) in Relish

RL Activation and Antibacterial Immune Responses in Drosophila.;

DR Science 296:359-362(2002).

SQ EMBL; AF500096; AAM18530.1; -

SQ SEQUENCE 500 AA; 54094 MW; A0459DE723A8C720 CRC64;

Query Match

Best Local Similarity 33.6%; Score 353.5; DB 5; Length 500;

Matches 67; Conservative 22; Mismatches 52; Indels 1; Gaps 1;

OY 47 LPLRYVVSHTAGSSCNTPASCOOQARNOHYHMKTLGMCVDYGNFLIGEDGLYEGRGW 106

DB 356 LPLRYVVSHTAGSSCNTPASCOOQARNOHYHMKTLGMCVDYGNFLIGEDGLYEGRGW 415

RESULT 14

O96PD5

ID O96PD5 PRELIMINARY; PRT; 576 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)

DE Peptidoglycan recognition protein L precursor.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-21433985; PubMed-11461926;

RT Peptidoglycan recognition proteins: a novel family of four human

RT innate immunity pattern recognition molecules.;

RL J. Biol. Chem. 276:34686-34694(2001).

DR EMBL; AF384856; AAL05629.1; -

KM Signal.

FT SIGNAL. 1 21 POTENTIAL.

SQ SEQUENCE 576 AA; 62216 MW; 73EA8713DC54F85A CRC64;

Query Match

Best Local Similarity 33.2%; Score 348.5; DB 4; Length 576;

Matches 70; Conservative 26; Mismatches 66; Indels 7; Gaps 5;

OY 25 CSPVPRNEMKALASE-CAOHLSPRLRYVVSHT--AGSSCNTPASCOOQARNOHYHMK 81

DB 379 CPAIHPRCRMGAPRYGRPRKLPQLPLGFLYHHTYVPAPCTDPTRCANRRSMQRYHQD 438

OY 82 TLGMCVDYGNFLIGEDGLYEGRGMTFGAHS-GHLMNPKSIGSPMGNTYDRVTPQAI 140

DB 439 TQGGDYGISFVSSDXYVEGGMHWGATLGH--NSRGEVAIVGNTAALPTFAAL 496

OY 141 RAAOGLL-ACGVAAGALNSNVLKGRDYORTLSPGNOLYHLIQNMPHY 188

DB 497 RYVADTLPSCAVRAGLPRDIALGHRQLVNTDPCGDALFDLKRMPHF 545

RESULT 15

O96N74

ID O96N74 PRELIMINARY; PRT; 634 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)

DE CNNA FLJ31320 f1s, clone LIVER100542, moderately similar to Mus

DE musculus TAGL-alpha mRNA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,

RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,

RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,

RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,

RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Maetsumura M., Marakami K., Kanehori K., Takahashi-Pujili A., Oshima A.
 RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
 RA Masuno Y., Nagai K., Itoai T.,
 RT "NEBD human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AK055882; BAB71034.1;
 SO SEQUENCE 634 AA; 67970 MW; 93E2032F3CE3BE70 CRC64;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 00:45:49 ; Search time 143 Seconds

(without alignments)
177.978 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051
Sequence: 1 MLAWALPRLRLGAQETE.....LSPGNOLYHLIQWPHYRSP 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002:.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1051	100.0	191	20	AA00771
2	1051	100.0	196	21	AA244022
3	1051	100.0	196	21	AA25583
4	1051	100.0	196	21	AA25583
5	1051	100.0	196	21	AA25583
6	1051	100.0	196	21	AA25583
7	1051	100.0	196	21	AA25583
8	652	65.8	182	20	AA00770
9	651.5	62.9	181	18	AA23723
10	494	47.0	116	21	AA64935

11	414.5	39.4	184	22	ABB69757	Drosophila melanog
12	413	39.3	241	22	ABB53271	Human polypeptide
13	413	39.3	369	22	AAE00693	Human full length
14	413	39.3	369	22	AAE00693	Human polypeptide
15	413	39.3	369	22	AAE00692	Human full length
16	405.5	38.6	185	22	ABB59234	Drosophila melanog
17	405.5	38.6	185	22	ABB69758	Drosophila melanog
18	379	36.1	203	22	ABB70267	Drosophila melanog
19	371.5	35.3	368	22	AA196963	Wound healing tiss
20	348.5	33.2	363	21	AA194863	Human protein clon
21	348.5	33.2	576	22	AA665916	Amino acid sequenc
22	348.5	33.2	634	22	AA665915	Amino acid sequenc
23	347.5	33.1	196	19	AAW37837	Amino acid sequenc
24	347.5	33.1	196	19	AAW37835	Amino acid sequenc
25	347	33.0	345	22	ABB64149	Drosophila melanog
26	332.5	31.6	530	22	AA172664	Murine peptidoglyc
27	330	31.4	186	22	ABB61200	Drosophila melanog
28	329.5	31.4	190	22	ABB64581	Drosophila melanog
29	328.5	31.3	173	19	AAW37834	Recombinant peptid
30	328.5	31.3	173	19	AAW37836	Amino acid sequenc
31	327	31.1	337	22	ABB60644	Drosophila melanog
32	317	30.2	215	22	ABB69711	Drosophila melanog
33	316.5	30.1	243	21	AA196962	C glutamyl prote
34	316.5	30.1	244	21	AA196962	Novel human diapo
35	289.5	27.5	132	22	AAE00694	Human peptidoglyc
36	248.5	23.6	611	22	ABB60610	Human zppai domain
37	229.5	21.8	280	22	ABB60590	Drosophila melanog
38	202	19.2	683	22	AA196963	Drosophila melanog
39	196.5	18.7	363	22	ABG27581	Human peptidoglyc
40	193	18.4	174	22	AA172663	Human peptidoglyc
41	177.5	16.9	132	22	AAE00694	Human zppai domain
42	147.5	14.0	505	22	ABB61601	Drosophila melanog
43	138	13.1	26	20	AA100774	Human tag7 clone p
44	127.5	12.1	33	22	ABB42485	Peptide #9991 enco
45	127.5	12.1	53	22	ABB25910	Protein #7909 enco

ALIGNMENTS

RESULT 1	AA00771	standard; Protein, 191 AA.
ID	AA00771	
XX	AA00771	
AC	AA00771	
DT	18-MAY-1999	(first entry)
XX	Human tag7 clone protein sequence.	
DE	Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma; melanoma; leukaemia; apoptosis inducer; human.	
KW	Homo sapiens.	
XX	WO9902686-A1.	
PN	21-JAN-1999.	
PD	10-JUL-1998;	98WO-EP04287.
XX	11-JUL-1997;	97US-0893764.
PR	(BOEH) BOEHRINGER INGELHEIM INT. GMBH.	
PA	Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;	
XX	WPI: 1999-120887/10.	
XX	N-PSDB: AAX21820.	
DR	New nucleic acid encoding tag7 - used to inhibit tumour growth and	
XX	induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and	
PT	leukaemia	

```

XX  Claim 23, Page 126-127, 138pp; English.
PS  This sequence is the human tag7 of the invention. Cells containing
XX  the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
CC  to produce and purify antibodies; to inhibit growth of mammalian tumours,
CC  especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
CC  prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
CC  and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
CC  osteo- or fibro-sarcoma), melanoma or leukaemia; and as a molecular
CC  weight marker. The tag7 polypeptide inhibits tumour growth and induces
CC  apoptosis. The tag7 coding sequences are also useful as probes for gene
CC  mapping and detection of tag7 gene expression, and as primers. Antibodies
CC  against tag7 are used as reagents for detecting tag7; as an antagonist of
CC  tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
CC  metastasis.
XX  Sequence 191 AA;
SQ
Query Match 100.0%; Score 1051; DB 20; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.7e-104;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAAALPSLLRGAAGETEDPACSPYPRNEMKALASECAQHSPLRYVVSHTAGS 60
DB 1 MLAAALPSLLRGAAGETEDPACSPYPRNEMKALASECAQHSPLRYVVSHTAGS 60
QY 61 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLYEGRGWNTGASHGLMNP 120
DB 61 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLYEGRGWNTGASHGLMNP 120
QY 121 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNYVLKGRHVDVORTLSPGNOLYH 180
DB 121 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNYVLKGRHVDVORTLSPGNOLYH 180
QY 181 LIQNMPHYRSP 191
DB 181 LIQNMPHYRSP 191

RESULT 2
AAB24022
ID AAB24022 standard; Protein; 196 AA.
XX
XX AAB24022;
AC
XX 25-JAN-2001 (first entry)
DT
XX
DE Human PRO1269 protein sequence SEQ ID NO:7.
XX
XX Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW identification; tumorigenesis; anticancer; detection.
XX
XX Homo sapiens.
OS
XX WO200053750-A1.
PN
XX 14-SEP-2000.
PD
XX
XX 02-DEC-1999; 99WO-US28551.
XX
XX 08-MAR-1999; 99WO-US05028.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
XX 29-OCT-1999; 99US-0162506.
XX
XX 30-NOV-1999; 99WO-US28313.
XX
XX 01-DEC-1999; 99WO-US28634.
XX
XX (GETH ) GENENTECH INC.
XX
XX Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
XX WPI: 2000-594320/56.
XX
XX N-PSDB; AAC58104.
DR

```

```

XX  Antibodies specific for PRO polypeptides; used to diagnose and inhibit
PT  the growth of tumors in mammals, and to identify inhibitors of PRO
PT  polypeptide activity or expression.
XX
XX Claim 61; Fig 4; 226pp; English.
XX
XX The present invention describes an antibody that binds to a human
CC  protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC  PRO334; PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
CC  PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC  anticancer activity and can be used to diagnose tumors in mammals, by
CC  detecting complex formation when the antibody is contacted with test
CC  cells. Increased expression of genes encoding (I) can also be detected
CC  to diagnose tumors. Agents which inhibit the activity of (I),
CC  especially the antibodies, or an antisense oligonucleotide which
CC  hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC  preferably by inducing cell death. Methods from the present invention
CC  can be used to identify compounds which inhibit the biological activity
CC  of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC  probes used in examples from the present invention for human PRO
CC  sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC  PRO polynucleotide and protein sequences given in the exemplification of
CC  the present invention.
XX
XX Sequence 196 AA;
SQ
Query Match 100.0%; Score 1051; DB 21; Length 196;
Best Local Similarity 100.0%; Pred. No. 2.7e-104;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAAALPSLLRGAAGETEDPACSPYPRNEMKALASECAQHSPLRYVVSHTAGS 60
DB 6 MLAAALPSLLRGAAGETEDPACSPYPRNEMKALASECAQHSPLRYVVSHTAGS 65
QY 61 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLYEGRGWNTGASHGLMNP 120
DB 66 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLYEGRGWNTGASHGLMNP 125
QY 121 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNYVLKGRHVDVORTLSPGNOLYH 180
DB 126 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNYVLKGRHVDVORTLSPGNOLYH 185
QY 181 LIQNMPHYRSP 191
DB 186 LIQNMPHYRSP 196

RESULT 3
AAB25583
ID AAB25583 standard; Protein; 196 AA.
XX
XX AAB25583;
AC
XX 21-NOV-2000 (first entry)
DT
XX
DE Htag7 protein encoded by human secreted protein gene #8.
XX
XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritis;
KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnerrary; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human.
XX
XX Homo sapiens.
OS
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX
XX

```


OY 1 MLTAMALPSLLRIGAAQETEDPACCSPIVPRNEKALASECAOHLSTPLRYVVSHTAGS 60
DB 6 MLTAMALPSLLRIGAAQETEDPACCSPIVPRNEKALASECAOHLSTPLRYVVSHTAGS 65
OY 61 SCFTPASCOOQAAHVQHYHMKTLGWCVDVGNFLIGEDGLVEGRGMFTGAHSGHLMNPM 120
DB 66 SCFTPASCOOQAAHVQHYHMKTLGWCVDVGNFLIGEDGLVEGRGMFTGAHSGHLMNPM 125
OY 121 STGISPMGNMNDVPPPOALRAAGGLLACGVAAGALASNYVLKGRVQRTLSFGNOLYH 180
DB 126 STGISPMGNMNDVPPPOALRAAGGLLACGVAAGALASNYVLKGRVQRTLSFGNOLYH 185
OY 181 LIONMPTHYRSP 191
DB 186 LIONMPTHYRSP 196
RESULT 5
AAV9400
ID AAV9400 standard; Protein; 196 AA.
XX AAV9400;
AC
DT 08-AUG-2000 (first entry)
XX
DE Human FRO1269 (UNQ639) amino acid sequence SEQ ID NO:216.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
XX transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
OS Homo sapiens.
XX
PN WO200012708-A2.
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.

PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 22-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 30-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 10-NOV-1998; 98US-0106934.
PR 17-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.

PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
XX WPI: 2000-237871/20.
XX N-PSDB; AAA37082.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
XX secreted PRO polypeptides, useful for screening of potential peptide or
XX small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12; Fig 122; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX receptor or secreted PRO polypeptides given in AA99340 to AA99462. The
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding them have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
XX PCR primers and hybridization probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
XX Sequence 196 AA;
XX
XX Query Match 100.0%; Score 1051; DB 21; Length 196;
XX Best Local Similarity 100.0%; Pred. No. 2.7e-104;
XX Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MLAMALPSLRLGAAOETEDPACCSPIVPRNEMKALASECAOHLSTPLRYVVSHTAGS 60
XX 6 MLAMALPSLRLGAAOETEDPACCSPIVPRNEMKALASECAOHLSTPLRYVVSHTAGS 65
XX
XX 61 SCNTPASCCOQOARNVQHYHMKTLGMCVDGYNFLIGEDGLVEGRGNFTGAHSGHLMNPM 120
XX 66 SCNTPASCCOQOARNVQHYHMKTLGMCVDGYNFLIGEDGLVEGRGNFTGAHSGHLMNPM 125
XX
XX 121 SIGISFMGNMYMDRVPTPOAIRAAGILACGVAOGALRSNYVLKGRDVOFTLSPGNQLYH 180
XX 126 SIGISFMGNMYMDRVPTPOAIRAAGILACGVAOGALRSNYVLKGRDVOFTLSPGNQLYH 185
XX
XX 181 LIONMPHYRSP 191
XX 186 LIONMPHYRSP 196
XX
XX Db
XX
XX RESULT 6
XX AAB6149
XX ID AAB6149 standard; protein: 196 AA.
XX
XX AC AAB6149;
XX
XX DT 02-APR-2001 (first entry)
XX
XX DE Protein of the invention #61.
XX
XX KW Secreted; transmembrane; gene therapy.

XX
XX Unidentified.
XX OS
XX WO200078961-A1.
XX
XX 28-DEC-2000.
XX
XX 18-FEB-2000; 2000WO-US04342.
XX
XX 23-JUN-1999; 99US-0141037.
XX 20-JUL-1999; 99US-0144758.
XX 26-JUL-1999; 99US-0145698.
XX 01-SEP-1999; 99WO-US20111.
XX 29-OCT-1999; 98US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28551.
XX 16-DEC-1999; 99WO-US30095.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
XX Watanabe CK, Williams PM, Wood WI;
XX
XX WPI: 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
XX useful as hybridization probes. In chromosome and gene mapping and gene
XX therapy -
XX
XX Claim 1; Fig 122; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
XX These proteins and the DNA encoding them may be used as hybridization
XX probes, in chromosome and gene mapping and in the generation of
XX anti-sense RNA and DNA. They may also be used to generate either
XX transgenic animals or knockout animals which are in turn useful for
XX development and screening of therapeutically useful reagents.
XX The nucleic acids may also be used in gene therapy.
XX
XX Sequence 196 AA;
XX
XX Query Match 100.0%; Score 1051; DB 22; Length 196;
XX Best Local Similarity 100.0%; Pred. No. 2.7e-104;
XX Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MLAMALPSLRLGAAOETEDPACCSPIVPRNEMKALASECAOHLSTPLRYVVSHTAGS 60
XX 6 MLAMALPSLRLGAAOETEDPACCSPIVPRNEMKALASECAOHLSTPLRYVVSHTAGS 65
XX
XX 61 SCNTPASCCOQOARNVQHYHMKTLGMCVDGYNFLIGEDGLVEGRGNFTGAHSGHLMNPM 120
XX 66 SCNTPASCCOQOARNVQHYHMKTLGMCVDGYNFLIGEDGLVEGRGNFTGAHSGHLMNPM 125
XX
XX 121 SIGISFMGNMYMDRVPTPOAIRAAGILACGVAOGALRSNYVLKGRDVOFTLSPGNQLYH 180
XX 126 SIGISFMGNMYMDRVPTPOAIRAAGILACGVAOGALRSNYVLKGRDVOFTLSPGNQLYH 185
XX
XX 181 LIONMPHYRSP 191
XX 186 LIONMPHYRSP 196
XX
XX Db
XX
XX RESULT 7
XX AAM23722
XX ID AAM23722 standard; protein: 190 AA.
XX
XX AC AAM23722;
XX
XX DT 18-FEB-1998 (first entry)

Query Match	62.9%	Score 661.5	DB 18	Length 181
Best Local Similarity	66.1%	Pred. No. 1.3e-62		
Matches 125	Conservative 20	Mismatches 35	Indels 9	Gaps 2

AA42265 to AA43075 represent novel 5' expressed sequence tag (EST) sequences, corresponding to human secreted proteins. AA564651 to CC AA55438 represent the EST-related proteins corresponding to AA42265 to CC AA43032. The 5' ESTs can be used for producing secreted human gene CC products. They can be used to identify and isolate 5' untranslated CC regions (UTRs) and upstream regulatory regions which control the CC location, development stage, rate, and quantity of protein synthetasis, as CC well as stability of mRNA. The ESTs are also useful as probes for CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can CC also be used in forensic procedures to identify individuals, or in CC diagnostic procedures to identify individuals having genetic diseases CC resulting from abnormal gene expression. The products may also be used CC in gene therapy protocols. The nucleic acids encoding signal peptides can be CC used for directing extracellular secretion of a polypeptide or the CC insertion of a polypeptide into a membrane, or importing a polypeptide CC into a cell. The proteins encoded by the EST sequences may be useful in CC treating a variety of human conditions. Secreted proteins have CC therapeutic value, and the identification of new secreted proteins is CC

CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 116 AA;

Query Match 47.08; Score 494; DB 21; Length 116;
Best Local Similarity 97.8%; Pred. No. 6; Le-45;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLAMALPSLRLGAOETDPACCSPIVPRNEMKALASCAOHLSPRLRVVSHRAGS 60
DB 6 MLAMALPSLRLGAOETDPACCSPIVPRNEMKALASCAOHLSPRLRVVSHRAGS 65

QY 61 SCNTPASCOQOARNVQHYHMKTLGMCVGYNFL 93
DB 66 SCNTXASCOQOARNVQHYHMKTLGMCVGYNFL 98

RESULT 11

ABB69757
ID ABB69757 standard; Protein; 184 AA.

AC ABB69757;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 36063.

KW Drosophila: developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL13860.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure: SEQ ID NO 36063; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins
XX (AB101840-AB116175).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 184 AA;

Query Match 39.4%; Score 414.5; DB 22; Length 184;
Best Local Similarity 47.2%; Pred. No. 3; Le-36;
Matches 77; Conservative 31; Mismatches 54; Indels 1; Gaps 1;

QY 28 IVPNEMKALASCAOHLSPRLRVVSHRAGSSCNTPASCOQOARNVQHYHMKTLGMD 87
DB 23 IISSEMGGRSATSKTSJANTISTAVIHHTAGNYSTAACTOIAOTIAYHMDSIGAMD 82

QY 88 VGNFLIGEDGLVYEGRGMTGASHGLIMPMISIGISFMGNVMDRVPPOAIRAAGLL 147
DB 83 IGVNFLIGDGNVYEGRGMTGASHGLIMPMISIGISFMGNVMDRVPPOAIRAAGLL 141

QY 148 ACGYAOGALRSNYYLKGHRDVOFTLSPGNOLYHLIONPHRS 190
DB 142 SDAVSRCQIVSGYLLYGHROYGSTECPGTINWNEIRTSNMKA 184

RESULT 12

ABB53271
ID ABB53271 standard; Protein; 241 AA.

AC ABB53271;

DT 12-FEB-2002 (first entry)

DE Human polypeptide #11.

KW Human; neurotrophic; neuroprotective; anticonvulsant; antidepressant;
neuroleptic; tranquilizer; antiarrhythmic; cardiant; antidiabetic;
antiinflammatory; antilipemic; hepatotropic; vitruicid; antidiabetic;
neurotrophic; anorectic; cytotatic; respiratory disease; liver disease;
cardiovascular disease; respiratory disease; liver disease;
renal disease; skeletal muscle disease; gastrointestinal disease;
placental disease; testicular cancer; male fertility; pancreatic disease.

OS Homo sapiens.

PN WO200181363-A1.

PD 01-NOV-2001.

PF 26-APR-2001; 2001WO-US13360.

PR 27-APR-2000; 2000US-199963P.

PR 11-MAY-2000; 2000US-203336P.

PR 25-MAY-2000; 2000US-207087P.

PR 26-MAY-2000; 2000US-207546P.

PA (SMIR) SMITHKLINE BECHAM CORP.

PA (SMIR) SMITHKLINE BECHAM PLC.

PI Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;

PI Lai Y, Xie Q;

DR WPI; 2002-041392/05.

DR N-PSDB; ABA90336.

PT Novel polypeptides and polynucleotides useful as a vaccine for
PT preventing and treating diseases associated the polypeptide, e.g.
PT Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,
PT asthma, amesias -

PS Claim 1; Page 79; 116pp; English.

XX The invention relates to an isolated polypeptide comprising a 277, 480,
XX 583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
XX 844, 782, 262, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,
XX 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
XX given in the specification. The polypeptides, modulators of the
XX polypeptides and antibodies against the polypeptides are useful for
XX treating diseases such as neurological and psychiatric diseases
XX including Alzheimer's, paraspinauclear palsy, Huntington's disease,
XX myotonic dystrophy, anorexia and depression; cardiovascular diseases
XX including congestive heart failure, Hodgkin's disease and myocardial
XX infarction; respiratory diseases including asthma, chronic obstructive
XX pulmonary disease, cystic fibrosis and adult respiratory distress
XX syndrome; liver diseases including hypercholesterolaemia, cirrhosis,

CC viral and nonviral hepatitis, Type II diabetes mellitus, and impaired
CC glucose tolerance; renal disease including renal failure, acute tubular
CC necrosis and glomerulonephritis; skeletal muscle diseases including
CC Eulenburg's disease, hypoglycemia and obesity; gastrointestinal
CC diseases including myxoma congenita and intestinal obstruction; lymph
CC diseases including lymphagielectasia; diseases of placenta including
CC choriocarcinoma; diseases of testes including testicular cancer,
CC male reproductive diseases including low testosterone and male
CC infertility; and disease of pancreas including diabetic ketoacidosis,
CC Type I and 2 diabetes and obesity. The present sequence is a
CC polypeptide of the invention.

Sequence **241 AA;**

Query Match	39.3%;	Score 413;	DB 23;	Length 241;
Best Local Similarity	40.0%;	Pred. No. 7.9e-36;		
Matches	76;	Conservative	42;	Mismatches 66;
				Indels 6;
				Gaps 3;

```

QY 4 AMAVPSILR-----GGAODETDPACCSPIYVPRNMRKSLASCAQHSLLPRIVVSTAG 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 SYVQPLVLYKGNCCLAPRKSTSLKACACGVPRSVWGARETHCR-MTLPAKYGIITHTAG 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 60 SSCNTPASCQQAANVQHYHMKTLGMDGVGNFLIGEDLVYEGRGNFTGAHSHLMP 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 RTCNISEDCRLLVLDIOSFYIDRLKSDIGNLFVLDGAIYEGGVNNGS-STPEYD 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 120 MSIGISPMGNTMDRVPPTPOAIRAAGLLGACVAGQALRSNTVYLGKHDVQRTISPGNQLY 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 IALGTTEPGFTGTPIPNAAALEAODLQCAVMKGYLTPNLLVGHSDVARTLSPGQALY 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 180 HLIQNPVYR 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 NIISTWPRFK 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
AAE00693
ID AAE00693 standard; Protein; 369 AA

AC AAE00693;
XX
DT 02-JUL-2001 (first entry)
XX
XX
DE Human full length granulocyte peptide homolog zgp1 protein #2

KW Human; granulocyte peptide A; GP-A; Zepal; cytosolic; antiinflammatory
KW pulmonary; dermatological; anti-microbial; gastrointestinal disease;
KW vulnary; dental caries; periodontal disease; gene therapy; AIDS
KW Acquired Immune Deficiency Syndrome; epithelial wound; cystic fibrosis;
KW infection; urinary tract; respiratory; vaginal; lung; skin; cancer;
KW ovarian; rectal; chromosome 1.

OS Homo sapiens.

FH	Key	Location/Qualifiers
FH	Key	Location/Qualifiers

FT	label = signal_peptide
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	1
46	1
47	1
48	1
49	1
50	1
51	1
52	1
53	1
54	1
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1
88	1
89	1
90	1
91	1
92	1
93	1
94	1
95	1
96	1
97	1
98	1
99	1
100	1

```
FT /label= mature_full_length_zgpal_protein_#2
```

[illegible]

Domain	52..183	/1ahe] = Domain 1
FT		
FT		

ET	Region	84 : .90	/label = Immigrant estimate
ET	Region	84 : .90	/label = Immigrant estimate

FT	Region	92-97	note- this region is specifically claimed in claim 18
FT			

```

F1 /label= hyalophilic_region
FT 105..110
Region

```

FT	Region	117..122
----	--------	----------

44 14004 27 Oct 1944 + 01-09 + 00

FT	Region	150..155	/label= Hydrophilic_region
FT	Region	150..157	/label= Immunogenic-epitope
FT	Region	151..156	/label= Hydrophilic_region
FT	Region	152..157	/label= Hydrophilic_region
FT	Region	184..209	/label= Hydrophilic_region
FT	Region	190..202	/label= linker
FT	Region	205..223	/label= Immunogenic-epitope
FT	Region	210..340	/label= Immunogenic-epitope
FT	Domain	210..340	/note= "This region is specifically claimed in claim 18"
FT	Region	217..223	/label= Domain_2
FT	Region	239..249	/label= Immunogenic-epitope
FT	Region	261..267	/note= "This region is specifically claimed in claim 18"
FT	Region	291..297	/label= Immunogenic-epitope
FT	Region	346..354	/label= Immunogenic-epitope
FT	Region		/label= Immunogenic-epitope

PN WO200129224-A2

PD 26-APR-2001
xy

PF 20-OCT-2000; 2000WO-US29177.

PR 20-OCT-1999; 99US-0160712.
PR 13-THU-2000; 2000US-0318070

XX
BA (ZYMO) ZYMOGENETICS INC

XX.
PT Conklyn NC. Adler DA. F

XX
DB
WPT: 2001-290918/30

XX DK N-PSDB; AAAD040006;

PT New granulocyte peptide homolog, zgpai polypeptide, for research
PT applications, diagnosis and treatment of cancer, peritoneal,
PT gastrointestinal disease, urinary tract, skin and lung infections
XX
PS Claim 14: Page 109-110; 11app: English.

The present sequence is human full length granulocyte peptide (GP-1), homolog, Zgpai protein. Zgpai gene is located on human chromosome 1. Zgpai polypeptides are useful for producing antibodies which are useful for detecting cancer. Zgpai polypeptides having anti-microbial activity are useful for treating dental caries, periodontal disease, thrush, detecting cancer. Zgpai polypeptides having anti-microbial activity are useful for treating dental caries, periodontal disease, thrush, gastrointestinal disease, urinary tract infections, vaginal infections, respiratory infections, acquired immune deficiency syndrome (AIDS) and lung infections associated with cystic fibrosis and prevention of infection in skin and other epithelial wounds. Zgpai-cyclokin fusion proteins are useful for enhancing in vivo killing of target tissues (epithelial cancers, and more specifically lung, ovarian and rectal cancers). Zgpai polypeptides, fragments, fusion proteins or agonists are useful in in vitro studies of exogenous microorganism infections such as bacterial, viral or fungal infection and also to study epithelial cell defensin induction in cell culture. Zgpai antibodies, polynucleotides and polypeptides are useful for detection of zgpai polypeptide, mRNA or anti-zgpai antibodies, thus serving as markers for detecting genetic diseases or cancers. Zgpai sequences are useful as diagnostics in forensic DNA profiling and as probes or primers to clone 5' non-coding regions of a zgpai gene. Zgpai antibodies are useful for tagging cells

CC that express zgal, for screening expression libraries and as
 CC neutralizing antibodies or as antagonists to block zgal activity
 CC in vitro and in vivo. Zgal gene is also useful in gene therapy.
 XX

Sequence 369 AA;

Query Match 39.3%; Score 413; DB 22; Length 369;
 Best Local Similarity 40.0%; Pred. No. 1.4e-35;
 Matches 76; Conservative 42; Mismatches 66; Indels 6; Gaps 3;

QY 4 AWALPSLR----LGAOETEDPACCSPIVRNEMKALASECAQHLSPRLRYVVSHTAG 59
 DB 181 SYVPLLVKGENCLAPRKSTSLKACPGVPRSVWAGRETCPR-MTLPAKYGIITHAG 239
 QY 60 SSCNTPASCCQOQARNVQHYHKKTLGMCVDVGNFLGEGDLYVEGRGNFTGASHGLNMP 119
 DB 240 RTCNIDSECRLLVLDIOFYIDRLKSCDIGNFLVGOCALTEGVGMVQGS-STPGYDD 298
 QY 120 MSIGISFNGNMDVPPPOAIRAAQGLACGVAOAGALNSNYLKGHRDORTLSPGNOLY 179
 DB 299 IALGITFMGTFGTGIPNNAALAAQDLIOCAMVKGILTPNLLVGHSDVARTLSFGQALY 358
 QY 180 HLIONMPHYR 189
 DB 359 NIISTWPHFK 368

RESULT 14

ABB53272
 ID ABB53272 standard; Protein: 369 AA.

AC ABB53272;

DT 12-FEB-2002 (first entry)

XX Human polypeptide #12.

XX Human; nootropic; neuroprotective; anticonvulsant; antidepressant;
 KW neuroleptic; tranquilizer; antiarrhythmic; cardiant; antiasthmatic;
 KW antiinflammatory; antidiabetic; hepatotropic; virucide; antidiabetic;
 KW nephrotropic; anorectic; cytostatic; vaccine; neurological disease;
 KW cardiovascular disease; respiratory disease; liver disease;
 KW renal disease; skeletal muscle disease; gastrointestinal disease;
 KW placental disease; testicular cancer; male fertility; pancreatic disease.

XX Homo sapiens.

PN MO200181363-A1.

PD 01-NOV-2001.

PF 26-APR-2001; 2001WO-US13360.

PR 27-APR-2000; 2000US-199963P.

PR 11-MAY-2000; 2000US-20336P.

PR 25-MAY-2000; 2000US-207087P.

PR 26-MAY-2000; 2000US-207546P.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;

PI Lai Y, Xie Q;

DR WPI: 2002-041392/05.

DR N-PSDB; ABA90337.

XX Novel polypeptides and polynucleotides useful as a vaccine for
 PT preventing and treating diseases associated the polypeptide, e.g.
 PT Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,
 PT asthma, amnesia

XX The invention relates to an isolated polypeptide comprising a 277, 480,
 CC 583, 581, 628, 494, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
 CC 844, 782, 244, 471, 485, 286, 533, 490, 619, 490, 462, 255,
 CC 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
 CC given in the specification. The polypeptides, modulators of the
 CC polypeptides and antibodies against the polypeptides are useful for
 CC treating diseases and antidiabetic agents. The polypeptides are useful for
 CC including Alzheimer's, parasympathetic and psychiatric diseases
 CC myotonic dystrophy, anorexia and depression; cardiovascular diseases
 CC including congestive heart failure, Hodgkin's disease and myocardial
 CC infarction; respiratory diseases including asthma, chronic obstructive
 CC pulmonary disease, cystic fibrosis and adult respiratory distress
 CC syndrome; liver diseases including hypercholesterolemia, cirrhosis,
 CC viral and nonviral hepatitis, Type II diabetes mellitus, and impaired
 CC glucose tolerance; renal disease including renal failure, acute tubular
 CC necrosis and glomerulonephritis; skeletal muscle diseases including
 CC Eulenburg's disease, hypoglycaemia and obesity; gastrointestinal
 CC diseases including myotonia congenita and intestinal obstruction; lymph
 CC choriocarcinoma; diseases of testes including testicular cancer,
 CC male reproductive diseases including low testosterone and male
 CC infertility; and disease of pancreas including diabetic ketoacidosis,
 CC Type 1 and 2 diabetes and obesity. The present sequence is a
 CC polypeptide of the invention.

Sequence 369 AA;

Query Match 39.3%; Score 413; DB 23; Length 369;
 Best Local Similarity 40.0%; Pred. No. 1.4e-35;
 Matches 76; Conservative 42; Mismatches 66; Indels 6; Gaps 3;

QY 4 AWALPSLR----LGAOETEDPACCSPIVRNEMKALASECAQHLSPRLRYVVSHTAG 59
 DB 181 SYVPLLVKGENCLAPRKSTSLKACPGVPRSVWAGRETCPR-MTLPAKYGIITHAG 239
 QY 60 SSCNTPASCCQOQARNVQHYHKKTLGMCVDVGNFLGEGDLYVEGRGNFTGASHGLNMP 119
 DB 240 RTCNIDSECRLLVLDIOFYIDRLKSCDIGNFLVGOCALTEGVGMVQGS-STPGYDD 298
 QY 120 MSIGISFNGNMDVPPPOAIRAAQGLACGVAOAGALNSNYLKGHRDORTLSPGNOLY 179
 DB 299 IALGITFMGTFGTGIPNNAALAAQDLIOCAMVKGILTPNLLVGHSDVARTLSFGQALY 358
 QY 180 HLIONMPHYR 189
 DB 359 NIISTWPHFK 368

RESULT 15

AAE00692
 ID AAE00692 standard; Protein: 375 AA.

AC AAE00692;

DE 02-JUL-2001 (first entry)

XX Human full length granulocyte peptide homolog zgal protein #1.

XX Human; granulocyte peptide A; GP-A; zgal; cytostatic; antiinflammatory;
 KW vulnary; dermatological; anti-microbial; gastrointestinal disease;
 KW pulmonary; dental caries; periodontal disease; gene therapy; AIDS;
 KW Acquired Immune Deficiency Syndrome; epithelial wound; cystic fibrosis;
 KW infection; urinary tract; respiratory; vaginal; lung; skin; cancer;
 KW ovarian; rectal; chromosome 1.

XX Homo sapiens.

OS Key

PH Reptide

FT Protein

Location/Qualifiers
 1..17
 /label= Signal_peptide
 18..375
 /label= Mature_full_length_zgal_protein_#1

FT Region /note- "this region functions as an immunogenic epitope"
 FT 47..52 /label- Immunogenic_epitope
 FT /note- "this region is specifically claimed in claim 18"
 FT 58..189 /label- Domain_1
 FT Region /label- Hydrophilic_region
 FT 98..103 /label- Hydrophilic_region
 FT 123..128 /label- Hydrophilic_region
 FT 152..157 /label- Hydrophilic_region
 FT Region /label- Hydrophilic_region
 FT 156..161 /label- Hydrophilic_region
 FT Region /label- Hydrophilic_region
 FT 157..162 /label- Hydrophilic_region
 FT Region /label- Hydrophilic_region
 FT 158..163 /label- Hydrophilic_region
 FT Region /label- Hydrophilic_region
 FT 190..215 /label- Linker
 FT 216..346 /label- Domain_2
 FT Domain
 PN WC200129224-A2.
 PD 26-APR-2001.
 PF 20-OCT-2000; 2000WO-US29177.
 PR 20-OCT-1999; 99US-0160712.
 PR 12-JUL-2000; 2000US-0218070.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Conklin DC, Adler DA, Fox BA;
 XX
 DR WPI; 2001-290918/30.
 DR N-PSDB; AAD04004.
 XX
 PT New granulocyte peptide homolog, zgal polypeptide, for research
 PT applications, diagnosis and treatment of cancer, periodontal,
 PT gastrointestinal disease, urinary tract, skin and lung infections
 XX
 PS Claim 14; Page 105-106; 114pp: English.

CC The present sequence is human full length granulocyte peptide (GP-A)
 CC homolog, zgal protein. zgal gene is located on human chromosome 1.
 CC zgal polypeptides are useful for producing antibodies which are useful
 CC for detecting cancer. zgal polypeptides having anti-microbial activity
 CC are useful for treating dental caries, periodontal disease, thrush,
 CC detecting cancer. zgal polypeptides having anti-microbial activity are
 CC useful for treating dental caries, periodontal disease, thrush,
 CC gastrointestinal disease, urinary tract infections, vaginal infections,
 CC respiratory infections, acquired immune deficiency syndrome (AIDS) and
 CC lung infections associated with cystic fibrosis and prevention of
 CC infection in skin and other epithelial wounds. zgal-cytokine fusion
 CC proteins are useful for enhancing in vivo killing of target tissues
 CC (epithelial cancers, and more specifically lung, ovarian and rectal
 CC cancers). zgal polypeptides, fragments, fusion proteins or agonists are
 CC useful in in vitro studies of exogenous microorganism infections such as
 CC bacterial, viral or fungal infection and also to study epithelial cell
 CC defense induction in cell culture. zgal antibodies, polynucleotides and
 CC polypeptides are useful for detection of zgal polypeptide, mRNA or
 CC anti-zgal antibodies, thus serving as markers for detecting genetic
 CC diseases or cancers. zgal sequences are useful as diagnostics in
 CC forensic DNA profiling and as probes or primers to clone 5' non-coding
 CC regions of a zgal gene. zgal antibodies are useful for tagging cells
 CC that express zgal, for screening expression libraries and as
 CC neutralizing antibodies or as antagonists to block zgal activity
 CC in vitro and in vivo. zgal gene is also useful in gene therapy.

XX
 XX
 Sequence 375 AA;
 50

Query Match 39.3%; Score 413; DB 22; Length 375;
 Best Local Similarity 40.0%; Pred. No. 1,4e-35;
 Matches 76; Conservative 42; Mismatches 66; Indels 6; Gaps 3;

QY	4	AWALPSLR---	LGAQETEDPACCSPVPRNEWKALASECAQHLSPRLRYVVSHTAG	59
DB	187	SYVQPLVYGENCLAPROKTSILKACPGVPRSVWGAFREHCPR-WLPAAKGIILHTAG	245	
QY	60	SSCNTPASCQOQARVQVHIMAKTLGMCVGYNPLIGEDGLVEGRGMNFTGAHSGHKNP	119	
DB	246	RTCNISDECRLLVRDIQSEFYIDRLKSCDIGNPLVGQDAIYEGVGMVQGS-STPGYDD	304	
QY	120	MSIGISFMGNMYMDRVPPTQOAIRAOCILACGVAOCALRSNYVLKGRDVOFTLSPGNOLY	179	
DB	305	IALGTFMGFTGIRPNAALPAADLLICANVKGILTYRNYLLVGHSDVARTLSPQALY	364	
QY	180	HLIQWPHYR	189	
DB	365	NIITWPHRK	374	

Search completed: November 12, 2002, 02:22:14
 Job time : 144 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 02:18:04 ; Search time 80 seconds
(without alignments)
229.521 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051

Sequence: 1 MLAWALPILRLGAQETE.....LSPGNLYHLIQNPHYRSP 191

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224	21.3	539	2	G70520
2	148.5	14.1	308	2	T35303
3	127	12.1	151	2	S07506
4	120	11.4	151	1	MOBPA7
5	91	8.7	879	2	F96558
6	82	7.8	268	2	T24371
7	79.5	7.6	267	2	H71878
8	79.5	7.6	267	2	B64636
9	78	7.4	432	2	C85087
10	77.5	7.4	277	2	I52825
11	77.5	7.4	282	2	S50031
12	77.5	7.4	345	2	T48111
13	77.5	7.4	356	2	D65096
14	77.5	7.4	356	2	A88124
15	77.5	7.4	356	2	H85968
16	76.5	7.3	379	2	AD3390
17	76.5	7.3	786	2	A47546
18	76	7.2	229	2	T03405
19	76	7.2	296	2	T35345
20	75.5	7.2	1467	2	T23950
21	75	7.1	764	2	S47569
22	75	7.1	803	2	A71475
23	74.5	7.1	313	2	JE0328
24	74.5	7.1	366	2	JE0328
25	74.5	7.1	739	2	A83015
26	74.5	7.1	818	2	B75606
27	74.5	7.1	851	2	F84639
28	74	7.0	713	2	A35502
29	74	7.0	761	2	JC7821

30	74	7.0	1477	2	T18534	protein-tyrosine k
31	73.5	7.0	244	2	AC3533	l-fucose phospho
32	73.5	7.0	420	2	H75395	ABC transporter, p
33	73.5	7.0	439	2	A55520	YscW protein - Yer
34	73.5	7.0	439	2	T43585	Yops secretion pro
35	73.5	7.0	641	2	G85043	hypothetical prote
36	73.5	7.0	717	2	S31035	retrovirus-related
37	73.5	7.0	1159	2	T02866	hypothetical prote
38	73.5	7.0	1733	1	B45344	hypothetical prote
39	73.5	6.9	393	2	JE0180	probable nuclear a
40	73	6.9	533	2	D75220	phosphopentomutase
41	72.5	6.9	378	2	B55580	hypothetical prote
42	72.5	6.9	532	2	A72694	cardamemyl-phosphat
43	72.5	6.9	628	2	AC1469	hypothetical prote
44	72.5	6.9	566	2	H83037	transporter homolo
45	72	6.9	566	2	H83037	urease alpha subun

ALIGNMENTS

RESULT 1
G70520
probable csp protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70520
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtrold, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; M01D:9629387; PMID:9634230
A:Accession: G70520
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-539 <COI>
A:Cross-references: GB:297188; GB:A1123456; M1D:93261805; PIDN:CAB10019.1; PID:el300C
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: csp

Query Match 21.3%; Score 224; DB 2; Length 539;
Best Local Similarity 29.0%; Pred. No. 1.3e-13;
Matches 64; Conservative 31; Mismatches 82; Indels 44; Gaps 8;

DB 5 WALPILRLGAQETEDPACCSPIVPRNEM---KALASECAOHLSPLRVVSHTAGSS 61
183 WPPRAVYVMAAG---QPPA-----ISRAMEGADESILRCETPER-DGVRRAAVVHHTRGSN 233
OY 62 CHTPASCOQAAANVOHYHMKTLGWCMDVGNFLIGEDGLVYEGRWNT---GAHSGHLW 117
DB 234 DYSPELAGIVKAIYTYHSHKTLGWCMDIAYNALVDKYGQVFEASAGILTRPVEGFHTGG-F 292
OY 118 NMSGISPMGKMDVRVPPQAIIRAQGLACGYA-----G 134
DB 293 NNTWGVAMIGNFDVAPPIQIRTVGRLLGWLGMDDVDVPRNSVMDLQAGSSYTPPGG 352
OY 155 ALRSNVYLKGRDVOFTLSPGNLYHLION---WPHYRSP 191
DB 353 AIALRLPAITFTHRDVGTDCPGNAAYAVMDEIRIDIAHFNDP 393

RESULT 2
T35303
hypothetical protein SC5F7.1ac - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35303
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574

A:Accession: T35303
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-308 <SEE>
 A:Cross-references: EMBL:AL096872; PIDN:CA851271.1; GSPDB:GN00070; SCOPDB:SCSF7.14C
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOPDB:SCSF7.14C

Query Match 14.1%; Score 148.5; DB 2; Length 308;
 Best Local Similarity 28.2%; Pred. No. 1.2e-06;
 Matches 66; Conservative 19; Mismatches 82; Indels 67; Gaps 12;

DB 3 LAMAPSLRLGA-----AQETEDPACCSPTVPRNEMKALSEC 41
 66 LIGALPBLAVALFLCANGVERAADAEVPAATDRPAAPRDVPSVW---LGDA 122
 QY 42 AOHLSLEPLRY-----VVVSHT---AGSSC-NTPASCOOQANVQHYHMKTLGMCQDVGN 91
 DB 123 ARAOPAP-RYDDEVAVAFVHHHTDPNGYDCAADVPAL---LRGVYGGQTGAAMDIDIGYN 177
 QY 92 FLIGEDGLYEGRGWNE---TGHSGHLMNPMSTIGSFNGNYMDRVPPTQAT----- 140
 DB 178 FVYDRCGTYEGRAGGIDRPVTSMT-OGFNHRTTGIALGTAYAGVPVDELTDATAAV 236
 QY 141 -----RAOGLACG-----VAOGLRSNVYLKGRDVRQRTLSPGNOL 178
 DB 237 AAMKLGSTGTDPRAKVALVSSNGLSRYAAGTAMLPVAGHEDGYQTSCEPAL 290

RESULT 3
 S07506
 N:acetylmutamoyl-L-alanine amidase (EC 3.5.1.28) - phage T3
 N:Alternate names: gene 3.5 protein
 C:Species: phage T3
 C:Date: 30-Sep-1991 #sequence, revision 30-Sep-1991 #text_change 22-Jun-1999
 C:Accession: S07506
 R:Beck, P.J.; Gonzalez, S.; Ward, C.L.; Molinieux, I.J.
 J. Mol. Biol. 210, 687-701, 1989
 A:Title: Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.
 A:Reference number: S07500; M01D:90133923; PMID:2614843
 A:Accession: S07506
 A:Molecule type: DNA
 A:Residues: 1-151 <BEC>
 A:Cross-references: EMBL:X17255; NID:g15682; PIDN:CA85133.1; PID:g15695
 C:Genetics:
 A:Gene: 3.5
 C:Superfamily: phage T7 N-acetylmuramoyl-L-alanine amidase
 C:Keywords: hydrolase

Query Match 12.1%; Score 127; DB 2; Length 151;
 Best Local Similarity 32.4%; Pred. No. 6.2e-05;
 Matches 36; Conservative 18; Mismatches 47; Indels 10; Gaps 4;

QY 73 RNVOHYHMKTLGMCQDVGNFLIGEDGLYEGRGWNEFGHSHGLMNPMSIGSFNGNYMD 132
 DB 31 REIRQWH-KDQGMWDVYHFLIKRDGYVEAGRDELAVSHAKG-YNNHNSICLVGGIDD 88
 QY 133 RVP-----TPQAIRAAGCLACGVAAGALRSNVYLKGRDVRQRTLSPGNOL 178
 DB 89 KGRFDANFTPAQMOSLRSLVTLAK---YEGSVLRAHHVDVAPRACPSFDL 136

RESULT 4
 M0BPAT
 N:acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) - phage T7
 N:Alternate names: T7 lysozyme
 C:Species: phage T7
 C:Date: 01-Sep-1981 #sequence, revision 01-Sep-1981 #text_change 18-Jun-1999
 C:Accession: G94615; D92866; S42302; S43502; A01001
 R:Dunn, J.J.; Thompson, K.
 submitted to the Nucleic Acid sequence database, September 1982
 A:Reference number: A94615

A:Accession: G94615
 A:Molecule type: DNA
 A:Residues: 1-151 <D01>
 R:Dunn, J.J.; Studier, F.W.
 J. Mol. Biol. 148, 303-330, 1981
 A:Title: Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the
 A:Reference number: A92866; M01D:82078034; PMID:7310871
 A:Accession: D92866
 A:Molecule type: DNA
 A:Residues: 1-151 <D02>
 R:Dunn, J.J.; Studier, F.W.
 J. Mol. Biol. 166, 477-535, 1983
 A:Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7
 A:Reference number: S42283; M01D:83241725; PMID:6864790
 A:Accession: S42302
 A:Molecule type: DNA
 A:Residues: 1-38, 'E', '40-151' <DUN>
 A:Cross-references: EMBL:V01146
 R:Dunn, J.J.; Studier, F.W.
 submitted to the EMBL Data Library, October 1993
 A:Reference number: S43501
 A:Accession: S43502
 A:Molecule type: DNA
 A:Residues: 1-38, 'E', '40-118', 'V', '120-151' <DUN>
 A:Cross-references: EMBL:V01146; NID:g431187; PIDN:CA824403.1; PID:g431190
 A:Comment: This late gene protein is not required for lysis but plays an important ro
 lly bound. The functional significance of this is not clear.
 C:Genetics:
 A:Gene: 3.5
 A:Map position: 26.77-27.90
 C:Superfamily: phage T7 N-acetylmuramoyl-L-alanine amidase
 C:Keywords: hydrolase

Query Match 11.4%; Score 120; DB 1; Length 151;
 Best Local Similarity 31.9%; Pred. No. 0.00029;
 Matches 36; Conservative 19; Mismatches 44; Indels 14; Gaps 5;

QY 73 RNVOHYHMKTLGMCQDVGNFLIGEDGLYEGRGWNEFGHSHGLMNPMSIGSFNGNYMD 132
 DB 31 REIRQWH-KDQGMWDVYHFLIKRDGYVEAGRDELAVSHAKG-YNNHNSICLVGGIDD 88
 QY 133 RVP-----TPQAIRAAGCLACGVA--OGALRSNVYLKGRDVRQRTLSPGNOL 178
 DB 89 KGRFDANFTPAQMOSLRSLVTLAKYGA-----GLAHHVAVAKACPSFDL 136

RESULT 5
 F96558
 Probable protein kinase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence, revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F96558
 R:Theodoridis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 Nansen, N.F.; Hughes, B.; Hutzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; M01D:21016719; PMID:11130712
 A:Accession: F96558
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-879 <STO>
 A:Cross-references: GB:AE005173; NID:g9802791; PIDN:AA99960.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T14L22.12
 A:Map position: 1

Query Match 8.7%; Score 91; DB 2; Length 879;

Query Match	7.48;	Score 78;	DB 2;	Length 432,
Best Local Similarity	33.38;	Pred. No. 11;		

XX 02-DEC-1999; 99WO-US28551.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WT;
XX WPI; 2000-594320/56.
DR P-PSDB; AAB24022.
XX
XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of PRO
PT polypeptide activity or expression -
XX
XX Claim 50; Fig 3; 226pp; English.
XX
XX The present invention describes an antibody that binds to a human
CC Protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO334; PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC anticancer activity and can be used to diagnose tumors in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumors. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human PRO
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
XX SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
XX
XX Alignment Scores:
Pred. No.: 5.17e-100 Length: 697
Score: 1051.00 Matches: 191
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-462-625-4 (1-191) x AAC58104 (1-697)
QY 1 MetleuleuAlaITPAlaLeuProSerleuLeuArgleuGlyAlaAlaGlnIuThrglu 20
Db 41 ATGCGCTTGCTGGGCTCCGCCAGCTCTCCATCGATCGAGCGGCTAGAGACAGAA 100
QY 21 AsPProAlaCysCysSerProIleValProArgAsnGluITPLeuAlaLeuAlaSerIu 40
Db 101 GACCGGCTCTGCTGACGCCCATAGTGGCCCGAAGAGAGTGAAGGCCCTGGCATCAGAG 160
QY 41 CysAlaGlnHisLeuSerleuProleuArgTyrValValSerHisThrAlaGlySer 60
Db 161 TGGCGCCAGCAGACCTAGGCTGCTTACGCTATGTGTGATGCGACACGGCGGCGACAG 220
QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisIsthrHisMet 80
Db 221 AGCTGCAACACCCCGCTGCTGCGACAGAGCGCCGGAATGTGACATACACCATG 280
QY 81 LysThrLeuGlyITPLeuAspValGlyTyrAsnPhleuIleGlyGluAspGlyLeuVal 100
Db 281 AAGACACTGGGCTGTGCGACCTGGGCTACAACTCTCTATGTGAGAAACAGCGGCTCGTA 340
QY 101 TyrGluGlyArgGlyITPAsnPhThrGlyAlaHisSerGlyHisLeuITPAsnPromet 120

Db 341 TACGAGGCGCGGCTGGAACCTTACAGGGTCCCACTCAGGTCACTATGAGAACCCCATG 400
QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
Db 401 TCCATTGGCATACGCTTCTGCGCAACTACATGATCGGCTGCCCAACCCAGGCCATC 460
QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
Db 461 CGGCGAGCCAGGAGGTCTACTGCGCTGGCGGTGTGGCTCAGAGAGCCCTGAGGTCCAACTAT 520
QY 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
Db 521 GTGCTCAAGGACACCGGATGTGCAGCGTACACTCTCTCAGGCAACACGCTTCACAC 580
QY 181 LeuIleGlnAsnITPProHisITPTrArgSerPro 191
Db 581 CTCATTCAGAAATTGGCCACACTACCGCTCCCC 613
RESULT 2
AAA37082
ID AAA37082 standard; cDNA; 697 BP.
XX
AC AAA37082;
XX
XX 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (UNQ639) cDNA sequence SEQ ID NO:215.
XX
KM Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW Transmembrane; secretion; Immunoadhesion; pharmaceutical; screening;
ss.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
PF 01-SEP-1999; 99US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100664.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.

DB 122 ATAGTCCCCGAGACAGTGGAGGCGCTGTGCATCAGAGTGGCGCCGACGACCTAGGCTG 181
OY 209 CCCTTACGCTATGTGTGTATGTCACACAGCGGGGAGCAGCTGCACACCCCCGCTCG 268
DB 182 CCTTACGCTATGTGTGTATGTCACACAGCGGGGAGCAGCTGCACACCCCCGCTCG 241
OY 269 TGGCAGCAGAGGCGCGGATGTGACGACCTACACATGAAAGACATGGGCTGTGGAC 328
DB 242 TGGCAGCAGAGGCGCGGATGTGACGACCTACACATGAAAGACATGGGCTGTGGAC 301
OY 329 GTGGGCTACACTTCCGATTGGAGAGAGGCGCTGTATACGAGGCGCGTGGAGAC 388
DB 302 GTGGGCTACACTTCCGATTGGAGAGAGGCGCTGTATACGAGGCGCGTGGAGAC 361
OY 389 TTACAGGCGTCCGACATGAGCTACATTATGAAACCCATGTCCATGGCATGCTTATG 448
DB 362 TTACAGGCGTCCGACATGAGCTACATTATGAAACCCATGTCCATGGCATGCTTATG 421
OY 449 GGCACACTACATGATGGTGGGTCGCCACACCCGACATCCGGGACGCCAGGCTTACTG 508
DB 422 GGCACACTACATGATGGTGGGTCGCCACACCCGACATCCGGGACGCCAGGCTTACTG 481
OY 509 GCCTGGGTGGTGGTGCAGGAGCCGTCAGTCCATATGTGTCAAAGAGACACCGGAT 568
DB 482 GCCTGGGTGGTGGTGCAGGAGCCGTCAGTCCATATGTGTCAAAGAGACACCGGAT 541
OY 569 GTGAGGCTACACTTCTCCAGGACACAGCTTACACCTCATCCAGATTTGGCCACAC 628
DB 542 GTGAGGCTACACTTCTCCAGGACACAGCTTACACCTCATCCAGATTTGGCCACAC 601
OY 629 TACCGCTCCCGCTAGGCGCTGTGATCCGACCCCATTCCTCCCTCCATGGCCAAA 688
DB 602 TACCGCTCCCGCTAGGCGCTGTGATCCGACCCCATTCCTCCCTCCATGGCCAAA 661
OY 689 ACCCAGCTGTCTCTCTCCCAATAAGATG 718
DB 662 ACCCAGCTGTCTCTCTCCCAATAAGATG 691

RESULT 3
AAA37082
ID AAA37082 standard; cDNA; 697 BP.
XX
AC AAA37082;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (UN0639) cDNA sequence SEQ ID NO:215.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
ss.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 02-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.

PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100663.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103329.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105002.
PR 20-OCT-1998; 98US-0105007.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.

PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 30-OCT-1998; 98US-0108500.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 10-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 18-NOV-1998; 98US-0108825.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WT;
XX
XX WPI: 2000-237871/20.
XX
XX P-PSDB; AAY9400.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
XX secreted PRO polypeptides, useful for screening of potential peptide or
XX small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 2; Fig 121; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX receptor or secreted PRO polypeptides given in AAY9340 to AAY93462. The
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding then have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
XX PCR primers and hybridisation probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
XX Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
XX
XX

Query Match 93.9%; Score 674; DB 21; Length 697;
Best Local Similarity 98.6%; Pred. No. 28-161;
Matches 680; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 29 CCTGACATGATCTGACGACCACTATGTCGCCGCTTATGCTGCTGCTGAGGCTTC 88
DB 2 CCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 61
OY 89 CCGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 148

DB 62 CCGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 121
OY 149 ATAGTGGCCCGGAGACAGTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 208
DB 122 ATAGTGGCCCGGAGAGTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
OY 209 CCGTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 268
DB 182 CCGTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 241
OY 269 TGGCAGCAGCAGGCGCGGAGTGGCAGCAGTGGCAGCAGTGGCAGCAGTGGCAGC 328
DB 242 TGGCAGCAGCAGGCGCGGAGTGGCAGCAGTGGCAGCAGTGGCAGCAGTGGCAGC 301
OY 329 GTGGGCTACACTTCTGATTTGGAAGACGCGCTGCTATGACGAGGCGCTGCTGAGC 388
DB 302 GTGGGCTACACTTCTGATTTGGAAGACGCGCTGCTATGACGAGGCGCTGCTGAGC 361
OY 389 TTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 448
DB 362 TTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
OY 449 GGCACACTACATGATGCGGTGCGCACACCCAGGCGCATCCGAGGCGCTGCTACTG 508
DB 422 GGCACACTACATGATGCGGTGCGCACACCCAGGCGCATCCGAGGCGCTGCTACTG 481
OY 509 GCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 568
DB 482 GCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541
OY 569 GTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 628
DB 542 GTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
OY 629 TACCGTCCCGCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 688
DB 602 TACCGTCCCGCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661
OY 689 ACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
DB 662 ACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 691
RESULT 4
AAF54356
ID AAF54356 standard; DNA; 697 BP.
XX
XX AAF54356;
XX
XX 02-APR-2001 (first entry)
XX
XX
XX DNA encoding protein of the invention #61.
XX
XX Secreted; transmembrane; gene therapy; ss.
XX
XX Unidentified.
XX
XX WO200078961-A1.
XX
XX
XX 28-DEC-2000.
XX
XX 18-FEB-2000; 2000WO-US04342.
XX
XX
XX 23-JUN-1999; 99US-0141037.
XX
XX 20-JUL-1999; 99US-0144758.
XX
XX 26-JUL-1999; 99US-0145698.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
XX 29-OCT-1999; 99US-0162506.
XX
XX 30-NOV-1999; 99WO-US28313.
XX
XX 02-DEC-1999; 99WO-US28551.
XX
XX 16-DEC-1999; 99WO-US30095.
XX
XX 05-JAN-2000; 2000WO-US00219.

PR 06-JUN-2000; 2000MO-US00376.
 PA (GENTR) GENENTECH INC.
 XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S,
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D,
 PI Watanabe CK, Williams PM, Wood WI.
 DR WPI: 2001-071395/08.
 XX
 PT Secreted and transmembrane proteins and nucleic acids designated PRO,
 PT useful as hybridization probes, in chromosome and gene mapping and gene
 therapy -
 XX
 PS Claim 2: Fig 121; 787pp; English.
 CC
 CC The present invention relates to secreted and transmembrane proteins.
 CC These proteins and the DNA encoding them may be used as hybridization
 CC probes, in chromosome and gene mapping and in the generation of
 CC anti-sense RNA and DNA. They may also be used to generate either
 CC transgenic animals or knockout animals which are in turn useful for
 CC development and screening of therapeutically useful reagents.
 CC The nucleic acids may also be used in gene therapy.
 XX
 SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
 Query Match 93.9%; Score 674; DB 22; Length 697;
 Best Local Similarity 98.6%; Pred. No. 2e-161;
 Matches 680; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 29 CCTGACATGTAACCTGACACCACTATGTCGCGGCTATGCTGCTGGCTGCTC 88
 DB 2 CCCGACCTGCGCGGCTGCACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTC 61
 QY 89 CCCAGCTCTCTGACTGAGACGCGCTCAGAGACAGAAAGCCGCTGCTCAGAGCCC 148
 DB 62 CCCAGCTCTCTGACTGAGACGCGCTCAGAGACAGAAAGCCGCTGCTCAGAGCCC 121
 QY 149 ATAGTCCCCCGAG 208
 DB 122 ATAGTCCCCCGAG 181
 QY 209 CCTTACGCTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 268
 DB 182 CCTTACGCTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
 QY 269 TGCAG 328
 DB 242 TGCAG 301
 QY 329 GTGGGCTACAACTTCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 388
 DB 302 GTGGGCTACAACTTCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 361
 QY 389 TTCACGGGTCGCCACTGAGTCACTTATGAAACCATGTCATGATGATGATGATGAT 448
 DB 362 TTCACGGGTCGCCACTGAGTCACTTATGAAACCATGTCATGATGATGATGATGAT 421
 QY 449 GGAACACTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 DB 422 GGAACACTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481
 QY 509 GCCTGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 568
 DB 482 GCCTGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 541
 QY 569 GTGACAGGTAACCTTCTCCAGGCAACACCTTACACACTGATCCAGATGATGATGAT 628
 DB 542 GTGACAGGTAACCTTCTCCAGGCAACACCTTACACACTGATCCAGATGATGATGAT 601
 QY 629 TACCCCTCCCTGAGGCGCTGCTGATCCGACACCCCATTCCTCCCTCCATGCCCAAA 688

DB 602 TACCGTCCCTCCGAGGCGCTGCTGATCGGACCCCATCTCTCCCTCCATGCCCAAA 661
 QY 689 ACCCAGCTCTCTCTCTCCCAATAAGATG 718
 DB 662 ACCCAGCTCTCTCTCTCCCAATAAGATG 691
 RESULT 5
 ID AAA80613 standard; cDNA; 726 BP.
 XX
 AC AAA80613;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Human Htag7 secreted protein gene #8.
 XX
 KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 KW antineumatic; dermatological; antiproliferative; antiarteriosclerotic;
 KW anticancer; vulnery; antiviral; antibacterial; antifungal;
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KW Crohn's disease; nephritis; hyperproliferative disorder;
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KW melanoma; lymphoma; wound healing; human; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200029435-A1.
 XX
 PD 25-MAY-2000.
 XX
 PF 27-OCT-1999; 99MO-US25031.
 XX
 PR 28-OCT-1998; 98US-0105971.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
 PI Greene JM;
 XX
 DR WPI: 2000-387742/33.
 DR P-PSDB: AAB25583.
 XX
 PT Isolated nucleic acid molecules encoding human secreted proteins are
 PT used for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PT wounds, and infectious diseases -
 XX
 PS Claim 1; Figure 34; 803pp; English.
 XX
 CC The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given
 CC in AAA80606-198623 encode the 12 secreted protein sequences given in
 CC AAB25576-825593. The human secreted proteins have various activities
 CC dependent on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant;
 CC anti-inflammatory; antiarthritic; antineumatic; dermatological;
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnery;
 CC antiviral; antibacterial; and antifungal activity. The proteins,
 CC polypeptides, agonists and antagonists may be used to treat prevent
 CC and/or diagnose various disease, disorders and conditions examples of
 CC which include: immune disorders e.g. Addison's disease, rheumatoid
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 CC hyperproliferative disorders such as paraptotinemias and purpura;
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 CC sequences may also be used in wound healing and the treatment of
 CC infectious diseases. The human secreted protein gene #8 and protein
 CC sequences are represented in sequences AAA80613 and AAB25583. Sequences
 CC AAA80662-A80663 represent genes related to the secreted protein gene#8.

Sequence 726 BP; 161 A; 251 C; 188 G; 126 T; 0 other;

Query Match 93.7%; Score 673; DB 21; Length 726;
Best Local Similarity 98.5%; Pred. No. 3,7e-161;
Matches 679; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

QY 30 CTGACATGTAACCTGACACCCATATGTCCTCCCTCTATGCTGTTCCCTGGCTCC 89
DB 9 CCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 68
QY 90 CCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 149
DB 69 CCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 128
QY 150 TAGTGGCCCCGGAAGAGTGAAGGCTGTCATGAGTGGGCCAGACCTGAGCTGTC 209
DB 129 TAGTGGCCCCGGAAGAGTGAAGGCTGTCATGAGTGGGCCAGACCTGAGCTGTC 188
QY 210 CCTTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269
DB 189 CCTTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 248
QY 270 GCGACGACAGGCGCGGATGTCAGCACTACACATGAAAGACCTGGCTGCTGCTGCTG 329
DB 249 GCGACGACAGGCGCGGATGTCAGCACTACACATGAAAGACCTGGCTGCTGCTGCTG 308
QY 330 TGGGCTACAACTCTCTGATGAGAGACGGGCTGATATGAGAGGCGCTGCTGCTGCTG 389
DB 309 TGGGCTACAACTCTCTGATGAGAGACGGGCTGATATGAGAGGCGCTGCTGCTGCTG 368
QY 380 TCACGGGTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 449
DB 369 TCACGGGTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 428
QY 450 GCAATACATGATGAGGAGGCGGACACCCAGGAGGCGGAGGCTGCTGCTGCTGCTG 509
DB 429 GCAATACATGATGAGGAGGCGGACACCCAGGAGGCGGAGGCTGCTGCTGCTGCTG 488
QY 510 CCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 569
DB 489 CCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 548
QY 570 TGCAGGCTACCTCTCTCCAGGACACCACTGCTACCACTGCTGCTGCTGCTGCTGCTG 629
DB 549 TGCAGGCTACCTCTCTCCAGGACACCACTGCTACCACTGCTGCTGCTGCTGCTGCTG 608
QY 630 ACCGCTCCCTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 689
DB 609 ACCGCTCCCTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 668
QY 690 CCCCACTGCTCTCTCTCCAAATTAAGATG 718
DB 669 CCCCACTGCTCTCTCTCCAAATTAAGATG 697

```

RESULT 6
AAAS1719
ID AAAS1719 standard; cDNA; 749 BP.

AAAS1719;
31-OCT-2000 (first entry)
Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.
peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator;
Chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
inhibitor; protein co-ordinate data; 89.
Homo sapiens.
Location/Qualifiers

```

FT CDS 55..645
FT /tag- a
FT sig_peptide 55..117
FT /tag- b
FT mat_peptide 118..642
FT /tag- c
PN W0200039327-A1.
XX
XX 06-JUL-2000.
XX
XX 22-DEC-1999; 99WO-US30736.
XX
XX 23-DEC-1998; 98US-0113809.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Young PE, Olsen HS;
XX WPI: 2000-452414/39.
XX P-PSDB; AAY96964.
XX
XX Claim 1; Fig 3; 191pp; English.
XX
XX Novel human peptidoglycan recognition protein-like proteins (PGRP)
XX expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
XX or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
XX proteins are useful for preventing, treating or ameliorating a medical
XX condition in a mammal (claimed). PGRP is useful in augmenting the immune
XX system in such areas as immune recognition, antigen presentation and
XX immune system activation. Antibodies or antagonists directed against
XX these proteins may be useful in reducing or eliminating disorders
XX associated with tumour necrosis factor (TNF) and TNF-like cytokines,
XX such as endotoxic shock and autoimmune disorders and for treating
XX infectious diseases including silicosis, sarcoidosis and idiopathic
XX pulmonary fibrosis.
XX
SQ Sequence 749 BP; 165 A; 259 C; 195 G; 130 T; 0 other;
Query Match 93.7%; Score 673; DB 21; Length 749;
Best Local Similarity 98.5%; Pred. No. 3,7e-161;
Matches 679; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 30 CTGACATGTAACCTGACACCCATATGTCCTCCCTCTATGCTGCTTGGCTGGCTCTCC 89
DB 32 CCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91
QY 90 CCAAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 149
DB 92 CCAAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 151
QY 150 TAGTGGCCCCGGAAGAGTGAAGGCTGTCATGAGTGGGCCAGACCTGAGCTGTC 209
DB 152 TAGTGGCCCCGGAAGAGTGAAGGCTGTCATGAGTGGGCCAGACCTGAGCTGTC 211
QY 210 CCTTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269
DB 212 CCTTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 271
QY 270 GCGACGACAGGCGCGGATGTCAGCACTACACATGAAAGACCTGGCTGCTGCTGCTG 329
DB 272 GCGACGACAGGCGCGGATGTCAGCACTACACATGAAAGACCTGGCTGCTGCTGCTG 331
QY 330 TGGGCTACAACTCTCTATGAGAGAGGCGCTGATGAGAGGCGCTGCTGCTGCTGCTG 389
DB 332 TGGGCTACAACTCTCTATGAGAGAGGCGCTGATGAGAGGCGCTGCTGCTGCTGCTG 391
QY 390 TCACGGGTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 449

```

PT Leukemia

XX Claim 11; Page 126-127; 138pp; English.

CC This sequence encodes the human tag7 of the invention. Cells containing
CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
CC to produce and purify antibodies; to inhibit growth of mammalian tumors,
CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular
CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
CC apoptosis. The tag7 coding sequences are also useful as probes for gene
CC mapping and detection of tag7 gene expression, and as primers. Antibodies
CC against tag7 are used as reagents for detecting tag7; as an antagonist of
CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
CC metastasis.

XX Sequence 718 BP; 142 A; 251 C; 194 G; 131 T; 0 other;

Query Match 100.0%; Score 718; DB 20; Length 718;
Best Local Similarity 100.0%; Pred. No. 1.4e-172;
Matches 718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGTTACTGGGCGCCAGAGGCTGGGCCCCGTGACATGTACTGACAGCCACTATGTCGG 60
DB 1 CTGAGTTACTGGGCGCCAGAGGCTGGGCCCCGTGACATGTACTGACAGCCACTATGTCGG 60
QY 61 CCAGCTGATGCTGCTTGGCCGCTGCTCCGACGCTCTCCGACGCTCTGACGAGCGGCTCAGGA 120
DB 61 CCAGCTGATGCTGCTTGGCCGCTGCTCCGACGCTCTCCGACGCTCTGACGAGCGGCTCAGGA 120
QY 121 GACAGAGAGCCGCGCTGCTGTCAGACCCCATATGTCGCCCGAGAGAGTGAAGGCGCTGAGC 180
DB 121 GACAGAGAGCCGCGCTGCTGTCAGACCCCATATGTCGCCCGAGAGAGTGAAGGCGCTGAGC 180
QY 181 ATGAGAGTGGCGCCAGACACTGAGCCTGACCTTACGCTATGTTGGTGTATGACACAGCGC 240
DB 181 ATGAGAGTGGCGCCAGACACTGAGCCTGACCTTACGCTATGTTGGTGTATGACACAGCGC 240
QY 241 GGGCAGCAGCTGCAACACCCCGCTGCTGTCAGACGAGCGCGGGAATGAGCACTA 300
DB 241 GGGCAGCAGCTGCAACACCCCGCTGCTGTCAGACGAGCGCGGGAATGAGCACTA 300
QY 301 CCACATGAAGACACTGGGCTGCTGTCAGACGAGCGGCTCACTTCTGATTTGAGAGACG 360
DB 301 CCACATGAAGACACTGGGCTGCTGTCAGACGAGCGGCTCACTTCTGATTTGAGAGACG 360
QY 361 GGTGATATGAGAGGCGGCTGCTGTCAGACGAGCGGCTCACTTCTGATTTGAGAGACG 420
DB 361 GGTGATATGAGAGGCGGCTGCTGTCAGACGAGCGGCTCACTTCTGATTTGAGAGACG 420
QY 421 CCCCATGTCATTTGGCATAGCTTCAATGAGCACTACATGATGGGTGCCACACCCCA 480
DB 421 CCCCATGTCATTTGGCATAGCTTCAATGAGCACTACATGATGGGTGCCACACCCCA 480
QY 481 GGGCATCCGGGAGCGCCAGAGGCTTACTGGCTGCGGTGTCAGAGGAGCCCTGAGGTC 540
DB 481 GGGCATCCGGGAGCGCCAGAGGCTTACTGGCTGCGGTGTCAGAGGAGCCCTGAGGTC 540
QY 541 CAACATATGTCCTCAAGAGACCGGGATGTCAGGCTACACTCTCCAGGCAACCAAGCT 600
DB 541 CAACATATGTCCTCAAGAGACCGGGATGTCAGGCTACACTCTCCAGGCAACCAAGCT 600
QY 601 CTACACATCTATCAGAAATTTGGCACACTACCGCTCCCTGAGGCGCTGATCGCA 660
DB 601 CTACACATCTATCAGAAATTTGGCACACTACCGCTCCCTGAGGCGCTGATCGCA 660
QY 661 CCCCATTTCTCCCTCCCTCCAGTGGCCAAAACCCCACTGTCTCTTCTCAATAAAGATG 718
DB 661 CCCCATTTCTCCCTCCCTCCAGTGGCCAAAACCCCACTGTCTCTTCTCAATAAAGATG 718

RESULT 2

AAC58104
ID AAC58104 standard; cDNA: 697 BP.

XX AAC58104;

DT 25-JAN-2001 (first entry)

DE Human PRO1269 nucleotide sequence SEQ ID NO:6.

KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KM identification; tumourigenesis; anticancer; detection; ss.

OS Homo sapiens.

PN W0200053750-A1.

PD 14-SEP-2000.

PE 02-DEC-1999; 99WO-US28551.

PR 08-MAR-1999; 99WO-US05028.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 01-DEC-1999; 99WO-US28654.

PA (GENT) GENENTECH INC.

PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;

DR WPI: 2000-594320/56.

DR P-PADB; AAB24022.

PT Antibodies specific for PRO polypeptides, used to diagnose and inhibit

PT the growth of tumors in mammals, and to identify inhibitors of PRO

PT polypeptide activity or expression

XX Claim 50; Fig 3; 226pp; English.

CC The present invention describes an antibody that binds to a human
CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO3434; PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC anticancer activity and can be used to diagnose tumours in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumours. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (I). AAC58109 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human PRO
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.

XX Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;

Query Match 93.9%; Score 674; DB 21; Length 697;
Best Local Similarity 98.6%; Pred. No. 2e-161;
Matches 680; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 29 CCTGAGCATGTACTGAGCAGCACTATGTCCTCCGCGCTATGCTGCTTGGGCTCTC 88
DB 2 CCGGAGCCTCTGCGCCCTGCACTATGTCCTCCGCGCTATGCTGCTTGGGCTCTC 61
QY 89 CCGAGCCTCTTGCATCGGAGCGGCTCAGAGAGACAGAACCCGGGCTCTGAGGCCC 148
DB 62 CCGAGCCTCTTGCATCGGAGCGGCTCAGAGAGACAGAACCCGGGCTCTGAGGCCC 121
QY 149 ATATGTCCTCCGAGAGAGTGAAGGCGCTGGCATGAGAGTGCAGCAGCACTGAGCCTG 208

Db 122 ATAGTCCCGGAGAGAGTGAAGGCCCTGGCATGAGATGGCCCGACGACCTGACCTG 181
QY 209 CCCCTACCGCTAGTGTGTGTATCGACACAGCGGGGAGCAGCTGCACACCCCGGCTCG 268
Db 182 CCCCTACCGCTAGTGTGTGTATCGACACAGCGGGGAGCAGCTGCACACCCCGGCTCG 241
QY 269 TGCACGACGACGAGCGCGGATGTGCAGCCTACACATGAAACACTGGGGCTGGTCCGAC 328
Db 242 TGCACGACGACGAGCGCGGATGTGCAGCCTACACATGAAACACTGGGGCTGGTCCGAC 301
QY 329 GTGGGCTACAACTTCCTGATTGGAGAGACGGGCTGTTAGAGAGGCGGCTGGTCCGAC 388
Db 302 GTGGGCTACAACTTCCTGATTGGAGAGACGGGCTGTTAGAGAGGCGGCTGGTCCGAC 361
QY 389 TTTCAGGGGTGGCCCACTGAGTACTTATGGAACCCCATGTCCATTGGCATYCAAGCTTCATG 448
Db 362 TTTCAGGGGTGGCCCACTGAGTACTTATGGAACCCCATGTCCATTGGCATYCAAGCTTCATG 421
QY 449 GGCACATCATGATCGGGTGGCCACACCCCGGATCGGGGAGCCCGAGGGTCTACTG 508
Db 422 GGCACATCATGATCGGGTGGCCACACCCCGGATCGGGGAGCCCGAGGGTCTACTG 481
QY 509 GCGTGGGCTGTGGCTCAGAGGAGCCCTGAGTCCACTATGTCTCAAAAGACACGGGAT 568
Db 482 GCGTGGGCTGTGGCTCAGAGGAGCCCTGAGTCCACTATGTCTCAAAAGACACGGGAT 541
QY 569 GTGAGGGTATACACTCTCTCCAGGACACCACTCTACCACTTCATCCAGATTGGCCACAC 628
Db 542 GTGAGGGTATACACTCTCTCCAGGACACCACTCTACCACTTCATCCAGATTGGCCACAC 601
QY 629 TACGGCTCCCGCTGAGGCGCTGCTGATCCGACCCCATTCGCCCTCCCATGGCCAAA 688
Db 602 TACGGCTCCCGCTGAGGCGCTGCTGATCCGACCCCATTCGCCCTCCCATGGCCAAA 661
QY 689 ACCCCACTGTCTCTCTCTCCATTAAGATG 718
Db 662 ACCCCACTGTCTCTCTCTCCATTAAGATG 691

RESULT 3
AAA37082
ID AAA37082 standard; cDNA; 697 BP.
XX
AC AAA37082;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (UNO639) cDNA sequence SEQ ID NO:215.
XX
KW Human: PRO polypeptide; membrane bound protein; receptor; diagnosis;
transmembrane; secretion; Immunoadhesion; pharmaceutical; screening;
ss.
XX
XX Homo sapiens.
OS
XX
PN WO200012708-A2.
PD 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20111.
PF
XX 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099556.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.

PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102309.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.

XX	27-OCT-1998;	98US-0105807.
PR	27-OCT-1998;	98US-0105881.
PR	27-OCT-1998;	98US-0105882.
PR	27-OCT-1998;	98US-0105882.
PR	28-OCT-1998;	98US-0106062.
PR	28-OCT-1998;	98US-0106023.
PR	28-OCT-1998;	98US-0106029.
PR	28-OCT-1998;	98US-0106030.
PR	28-OCT-1998;	98US-0106032.
PR	28-OCT-1998;	98US-0106033.
PR	28-OCT-1998;	98US-0106178.
PR	29-OCT-1998;	98US-0106248.
PR	29-OCT-1998;	98US-0106384.
PR	29-OCT-1998;	98US-0108500.
PR	30-OCT-1998;	98US-0106464.
PR	03-NOV-1998;	98US-0106856.
PR	03-NOV-1998;	98US-0106902.
PR	03-NOV-1998;	98US-0106905.
PR	03-NOV-1998;	98US-0106919.
PR	03-NOV-1998;	98US-0106932.
PR	03-NOV-1998;	98US-0106934.
PR	10-NOV-1998;	98US-0107783.
PR	17-NOV-1998;	98US-0108775.
PR	17-NOV-1998;	98US-0108779.
PR	17-NOV-1998;	98US-0108787.
PR	17-NOV-1998;	98US-0108788.
PR	17-NOV-1998;	98US-0108801.
PR	17-NOV-1998;	98US-0108802.
PR	17-NOV-1998;	98US-0108806.
PR	17-NOV-1998;	98US-0108807.
PR	18-NOV-1998;	98US-0108849.
PR	18-NOV-1998;	98US-0108848.
PR	18-NOV-1998;	98US-0108849.
PR	18-NOV-1998;	98US-0108850.
PR	18-NOV-1998;	98US-0108851.
PR	18-NOV-1998;	98US-0108852.
PR	18-NOV-1998;	98US-0108855.
PR	18-NOV-1998;	98US-0108856.
PR	18-NOV-1998;	98US-0108904.
XX	(GETH) GENENTECH INC.	
PA		
PI	Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;	
XX	WPI: 2000-237871/20.	
DR	P-PSDB; AAY99400.	
XX		
XX	New mammalian DNA sequences encoding transmembrane, receptor or	
PT	secreted PRO polypeptides, useful for screening of potential peptide or	
PT	small molecule inhibitors of the relevant receptor/ligand interactions	
XX		
PS	Claim 2; Fig 121; 773bp; English.	
XX		
CC	AAA37022 to AAA37144 encode the new isolated human transmembrane,	
CC	receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. Th	
CC	transmembrane and receptor PRO proteins can be used for screening of	
CC	potential peptide or small molecule inhibitors of the relevant	
CC	receptor/ligand interactions. The polypeptides and nucleotide sequences	
CC	encoding then have various industrial applications, including uses as	
CC	pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent	
CC	PCR primers and hybridisation probes used in the isolation of the PRO	
CC	polypeptides from the present invention.	
XX		
XX	Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;	

Query Match 93.99; Score 674; DB 21; Length 697;
 Best Local Similarity 98.66; Prod. No. 2e-161;
 Matches 680; Conservative 0; Mismatches 10; Indels 0; Gaps

Oy	29	CCTGGACATGTAACTGCGACGCCACATATGTCGCCGGCTTATAGTCGTGGCTTGGGCTCTC	88
Db	2	CCCGGACCCCTCCCGCCCTGCGCCACATATGTCGCCGGCTTATAGTCGTGGCTTGGGCTCTTC	61
Oy	89	CCGACGCTCTTGTGACTGCGAGCGGAGCTAGGAGACAGAGAACCCGGGCTTGTCTGAGAGCCCC	148

Db	62	CCACACCTCTTTCGACTGCGAGCGGCTCTGAGAGACAGAAACCCGGGCTGCTGCAGCCCC	121
QY	149	ATAGTGGCCCGGAAACGAGTGGAAAGGCCCTGGCATTCAGAGTCCGGCCAGCACTGACGCTG	208
Db	122	ATATGTGCCCGGAACGAATGGAAAGGCCCTGGCATTCAGAGTGGCCGCGCACCTGAGCCTG	181
QY	209	CCCTTACGCTATGTTGTTGATTCGCACAGCGGGGAGCAGACTGCAACACCCCGGCTCG	268
Db	182	CCCTTACGCTATGTTGTTGATTCGCACAGCGGGGAGCAGACTGCAACACCCCGGCTCG	241
QY	269	TGCCAGCAGCAGAGGCCCGGAATGTGCAGCAGTACCAACATGAAGACATCGGCTGGTGCAC	328
Db	242	TGCCAGCAGCAGAGGCCCGGAATGTGCAGCAGTACCAACATGAAGACATCGGCTGGTGCAC	301
QY	329	GTGGGCTACCAACTTCCTTATTTGGAGAAAGACGGGCTCGTATTCAGAGGCGCTGGTGCAC	388
Db	302	GTGGGCTACCAACTTCCTTATTTGGAGAAAGACGGGCTCGTATTCAGAGGCGCGTGGTGCAC	361
QY	389	TTACAGGGGTGCCATCTCAGCTCAGCTATATGAAACCCATGTCCATTTGGATCAGCTTATG	448
Db	362	TTTACAGGGGTGCCATCTCAGCTCAGCTATATGAAACCCATGTCCATTTGGATCAGCTTATG	422
QY	449	GGCAACTACATGATCGGGTGCACACACCCAGGCCATTCGGGGAGCCACAGGGTCTACTG	508
Db	422	GGCAACTACATGATCGGGTGCACACACCCAGGCCATTCGGGGAGCCACAGGGTCTACTG	481
QY	509	GCTTCGGGTGTGGCTCAGAGGAGCCCTGAGTTCCAACTATGTGCTCAAAAGACACCGGGAT	568
Db	482	GCTTCGGGTGTGGCTCAGAGGAGCCCTGAGTTCCAACTATGTGCTCAAAAGACACCGGGAT	541
QY	569	GTGCAAGCTTACACTCTCTCCAGGACACAGCTTACACCTCATCCAGATTTGGCCACAC	628
Db	542	GTGCAAGCTTACACTCTCTCCAGGACACAGCTTACACCTCATCCAGATTTGGCCACAC	601
QY	629	TACGCTCCCCCTGAGGCCCTGTGTATCCGACACCCCATTCCTCCCTCCATGGCCAAA	688
Db	602	TACGCTCCCCCTGAGGCCCTGTGTATCCGACACCCCATTCCTCCCTCCATGGCCAAA	661
QY	689	ACCCCACTGTCTCCTTCCTCCCAATAAAGATG	718
Db	662	ACCCCACTGTCTCCTTCCTCCCAATAAAGATG	691
-RESULT 4			
AAFS4356			
XX	AAFS4356 standard; DNA; 697 BP.		
XX	AAFS4356;		
XX	02-APR-2001 (first entry)		
DE	DNA encoding protein of the invention #61.		
XX	Secreted; transmembrane; gene therapy; ss.		
OS	Unidentified.		
XX	WO200078961-A1.		
XX	28-DEC-2000.		
XX	18-FEB-2000; 2000WO-US04342.		
XX	23-JUN-1999; 99US-0141037.		
PR	20-JUL-1999; 99US-0144758.		
PR	26-JUL-1999; 99US-0145698.		
PR	01-SEP-1999; 99WO-US02011.		
PR	29-OCT-1999; 99US-0162506.		
PR	30-NOV-1999; 99WO-US28313.		
PR	02-DEC-1999; 99WO-US28551.		
PR	16-DEC-1999; 99WO-US30095.		
PR	05-JAN-2000; 2000WO-US00219.		

RESULT 4	
AAFs4356	
ID	AAFs4356 standard; DNA; 697 BP.
XX	
AC	AAFs4356;
XX	
DT	02-APR-2001 (first entry)
XX	
DE	DNA encoding protein of the invention #61
XX	
KW	Secreted; transmembrane; gene therapy; ss
XX	
OS	Unidentified.
XX	
PN	WO200078961-A1.
XX	
PD	
XX	
XX	28-DEC-2000.
PF	
XX	18-FEB-2000; 2000MO-US04342.
XX	
PR	23-JUN-1999; 99US-0141037.
PR	20-JUL-1999; 99US-0144758.
PR	26-JUL-1999; 99US-0145698.
PR	01-SEP-1999; 99MO-US20111.
PR	29-OCT-1999; 99US-0162506.
PR	30-NOV-1999; 99MO-US28313.
PR	02-DEC-1999; 99MO-US28551.
PR	16-DEC-1999; 99MO-US30095.
PR	05-JAN-2000; 2000MO-US00219.


```

Db 392 TCACGGGGGCGCCAGCTACAGTCTATGAGACCCCATGCTTCATTTGGCATGACCTTCATG 451
OY 450 GCAACTACATGATGAGTGGGTGCGCCACACCCAGGCGATCCGGGACGCCAGGGTCTACG 509
Db 452 GCAACTACATGATGAGTGGGTGCGCCACACCCAGGCGATCCGGGACGCCAGGGTCTACG 511
OY 510 CCGTGGGTGGTGGTCCAGGAGACCCCTGAGGTCAACTATGCTCTCAAAAGACACCGGGATG 569
Db 512 CCGTGGGTGGTGGTCCAGGAGACCCCTGAGGTCAACTATGCTCTCAAAAGACACCGGGATG 571
OY 570 TGCAGCTACACTCTCTCCAGGACCAACAGCTCTACACCTATCATGATGGCCACT 629
Db 572 TGCAGCTACACTCTCTCCAGGACCAACAGCTCTACACCTATCATGATGGCCACT 631
OY 630 ACCGCTCCCGCTGAGGCGCTGATGATCGGACCCATTCCTCCCTCCCATGGCCAAA 689
Db 632 ACCGCTCCCGCTGAGGCGCTGATGATCGGACCCATTCCTCCCTCCCATGGCCAAA 691
OY 690 CCCAGCTGCTCTCTCTCCCAATAAAGATG 718
Db 692 CCCAGCTGCTCTCTCTCCCAATAAAGATG 720

```

RESULT 7

AA78509
ID AA78509 standard; cDNA; 688 BP.

AA78509;

18-FEB-1998 (first entry)

Bovine granulocyte peptide A precursor cDNA.

Antimicrobial peptide; antibacterial; antifungal;
fungicide; antiprotosoa; protozoacide; antiviral; virucide;
bovine granulocyte peptide A; BGP-A; preservative; sepsis;
endotoxaemia; cattle; ss.

Bos taurus.

Location/Qualifiers
29..601
/*tag= a
29..91
/*tag= b
560..598
/*tag= c
/product= BGP-A

MO9729765-A1.

21-AUG-1997.

13-FEB-1997; 97MO-US02218.

16-FEB-1996; 96US-0011834.

(REBC) UNIV CALIFORNIA.

Selsted ME;

WPI; 1997-424753/39.

P-PDB; AAM23722.

Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -
useful therapeutically, as preservatives for food, in water
treatment and in agriculture

Claim 13; Fig 4; 56pp; English.

This cDNA sequence encodes the precursor (see AAM23722) of a novel,
claimed antimicrobial peptide from bovine neutrophils, designated

CC bovine granulocyte peptide A or BGP-A (see AAM23722). It was
CC isolated from bovine bone marrow cDNA using PCR and RACE
CC procedures. The encoded BGP-A precursor comprises a 21-amino acid
CC signal peptide, a 156-residue propeptide and 13 C-terminal amino
CC acids corresponding to mature BGP-A. BGP-A and the murine
CC homologue, MCP-A (see AAM23725), exhibit activity against
CC Gram-positive and Gram-negative bacteria, fungi and viruses,
CC specifically *Staphylococcus aureus*, *Escherichia coli*, *Candida*
CC *albicans*, *Salmonella typhimurium* and *C. neoformans* (claimed).
CC They can be used in human or veterinary medicine (particularly to
CC treat disorders associated with lipopolysaccharides, e.g. sepsis
CC and endotoxaemia) or as preservatives in food products or in water
CC supplies (claimed). They can also be applied to crops to reduce
CC post-harvest spoilage or expressed in transgenic plants to increase
CC their disease resistance. They have low immunogenicity.

Sequence 688 BP; 126 A; 234 C; 201 G; 127 T; 0 other;

Query Match 48.7%; Score 349.4; DB 18; Length 688;

Best Local Similarity 72.9%; Pred. No. 4,5e-79; Matches 485; Conservative 0; Mismatches 161; Indels 19; Gaps 2;

```

OY 51 CTATGTCGGCGGCTGATGCTGCTGGCTGCTCCAGCCTCTTGACTGAG 110
Db 27 CCAATGTCGGCGGCTGATGCTGCTGGCTGCTCCAGCCTCTTGACTGAG 86
OY 111 CGGCTCAGAGACAGAAAGACCCGCTGCTGACGCCCATGATGATGGA 170
Db 87 CGGCTCAAGA-----CTGCGGAGATGCTGTCCCGGAAAGTGG 128
OY 171 AGGCTTGGCATGAGTGGCGCCAGCAGCTGAGCTGCTTACGCTATGATG 230
Db 129 GCGGCTGGCATGAGTGGCGCCAGGAGGCTGATGAGCGCTGCTGATG 188
OY 231 CGCAGAGGGGGGAGAGAGTGGCAACCCCGCTGCTGCGACAGAGCCGATG 290
Db 189 GCGACAGGGGGGAGAGGCTGCAACCTCGGCTGCTGCAAGGAGGAGCCAAAAG 248
OY 291 TGCAGCACTACACATGAAACACACTGGGCTGTCGAGCGGGGTAAACTCT 350
Db 249 TGCAGTACTACACAGTGGCGGAGCGGGGCTGTCGAGCGGGGTAAACTCT 308
OY 351 GAGAAGAGGGGCTGATACAGAGGCGGTGCTGAACCTTACAGGGTCCACT 410
Db 309 GAGAAGATGGGCTGCTGATGAGGCGGGGCTGGAACACTTATGCTCCTG 368
OY 411 ACTTATGAACCCCATGTCATGCGATGCTTATGCGCACTACATGATGATG 470
Db 369 CACGCTGAACCCCATGTCATGCGATGCTTATGCGCACTACATGATGATG 428
OY 471 CCACACCCCGGATCCGCGAGCGAGGCTTACGCGCGTGGTGGCTCAGGAG 530
Db 429 CCGCGGCTTCTGCTCAGAGCGCGGCGGAGAGTCTGCTGCTGCGGAGAT 488
OY 531 CCTGAGGTCATATGCTCAAAAGACACCGGATGTCAGCTTCTTCAG 590
Db 489 ACCTGACTCTTAACTAGAAAGTCAAAAGACACCGGATGTCAGAGAGCTCT 548
OY 591 GCAACAGCTTACACACCTCATCAGAAATGGCCACTACGCTCCCTGAGGCTG 650
Db 549 GGGAGAGCTCTATAAATCATCAGAGTGGCGGACATACGCGCGTGTAGGCGCT 608
OY 651 CTGATCCGACCCCAATTCCTCCGCTCCATGGCCAAAACCCCACTGCTCTT 710
Db 609 TCCGCTTC-TCACACCCCACTCCATCAGAAACCCCACTGCTTCCCTG 667
OY 711 TAAAG 715
Db 668 TAAAG 672

```

RESULT 8
AA242549

ID	AA42549	standard; cDNA; 380 BP.
XX	AA42549;	
AC	AA42549;	
DT	01-FEB-2000	(first entry)
XX		
DE	Human 5' EST isolated from a cDNA library SEQ ID NO:308.	
XX		
KW	Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;	
KW	gene therapy; chromosome mapping; upstream regulatory sequence;	
KW	forensic; location; development; protein synthesis; stability;	
KW	regulation; identification; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	MO9953051-A2.	
PD		
XX	21-OCT-1999.	
PF		
XX	09-APR-1999; 99MO-IB00712.	
XX		
PR	09-APR-1998; 98US-0057719.	
PR	28-APR-1998; 98US-0069047.	
XX		
PA	(GEST) GENSET.	
XX		
PI	Dumas Milne Edwards J, Duclert A, Giordano J;	
XX		
DR	WPI: 2000-038446/03.	
DR	P-PSDB; AAY64935.	
PT		
XX	Novel secreted protein 5' expressed sequence tag sequences used in	
XX	diagnostic, forensic, gene therapy, and chromosome mapping procedures	
PS	Claim 1; Page 336; 837pp; English.	
CC	AA42265 to AA42075 represent novel 5' expressed sequence tag (EST)	
CC	sequences, corresponding to human secreted proteins. AAY4651 to	
CC	AAY65438 represent the EST-related proteins corresponding to AA42265 to	
CC	AA43052. The 5' ESTs can be used for producing secreted human gene	
CC	products. They can be used to identify and isolate 5' untranslated	
CC	regions (UTRs) and upstream regulatory regions which control the	
CC	location, development stage, rate, and quantity of protein synthesis, as	
CC	well as stability of mRNA. The ESTs are also useful as probes for	
CC	chromosome mapping, and to obtain full length cDNA clones. The ESTs can	
CC	also be used in forensic procedures to identify individuals, or in	
CC	diagnostic procedures to identify individuals having genetic diseases	
CC	resulting from abnormal gene expression. The products may also be used in	
CC	gene therapy protocols. The nucleic acids encoding signal peptides can be	
CC	used for directing extracellular secretion of a polypeptide or the	
CC	insertion of a polypeptide into a membrane, or importing a polypeptide	
CC	into a cell. The proteins encoded by the EST sequences may be useful in	
CC	treating a variety of human conditions. Secreted proteins have	
CC	therapeutic value, and the identification of new secreted proteins is	
CC	valuable. AA42249 to AA42264 and AAY64644 to AAY64650 represent	
CC	sequences used in the exemplification of the present invention.	
XX		
XX		
SEQ	Sequence 380 BP; 67 A; 130 C; 114 G; 63 T; 6 other;	
Query Match	47.7%; Score 342.2; DB 21; Length 380;	
Best Local Similarity	95.0%; Pred. No. 2.6e-77;	
Matches 358; Conservative	5; Mismatches 13; Indels 1; Gaps 1	
0Y	24 GGGCCCCGAGATGATGACTGCAGCCACATATGCCGCCGCGCTATGATGCGCTGGG 83	
DB	4 GGTTCCTCGGGCCCTGCCCCCTGCACATATGTCGCCGCGCTATATGCTGCTGGG 63	
0Y	84 CTTCTCCCAAGCCTCTTTCGACTCGAGCGGCTCAGAGACAGAAAGACCCGCTGCTGCA 143	
DB	64 CTTCTCCCAAGCCTCTTTCGACTCGAGCGGCTCAGAGACAGAAAGACCCGCTGCTGCA 123	
0Y	144 GCCCATTAATGCCCCGGGAAAGAGTGGAGGCCCTGGCATCAAGTGCGCCACAGACTTGA 203	

Db	124	GCCCCATATGTCGCCCGGAAGAGTGGAAAGCCCTGGCATCAGATGCGCCCGACACTGA	183
QY	204	GCCTGCGCTTACGCTATATGTGTGATATGGACACACGCGCGGACGACAGCTCAACACCCCGG	263
Db	184	GCCTGCGCTTACGCTATATGTGTGATATGGACACACGCGCGGACGACAGCTCAACACCCCGG	243
QY	264	CCTGTGTCACAGCAGGCGCGGAAATGTGCAGCAGCTACACATGAAGACATGGGCTGT	323
Db	244	CCTGTGTCACAGCAGGCGCGGAAATGTGCAGCAGCTACACATGAAGACATGGGCTGT	303
QY	324	GCAGCTGGGGCTTACAACTTCTCT-GATTGGAGAAGACGGGCTGTATACGAGGCGCGTGC	382
Db	304	GCAGCTGGGGCTTACAACTTCTCTNGATTGGAGAAGACGGGCTGTATACAGGCGCGTGM	363
QY	383	TGGAACCTTCACGGGTGC	399
Db	364	TGGAACCTTCACGGGTGC	380
RESULT	9		
AXX21819			
ID	AXX21819	standard; cDNA; 549 BP.	
XX	AXX21819;		
AC			
XX	18-MAY-1999	(first entry)	
DT			
XX			
DE		Mouse tag7 clone coding sequence.	
XX			
Tag7;		tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;	
KM		melanoma; leukaemia; apoptosis inducer; mouse; ds.	
XX			
OS		Mus sp.	
WO9902686-A1.			
PN			
PD		21-JAN-1999.	
XX			
PF		10-JUL-1998; 98WO-EP04287.	
PR		11-JUL-1997; 97US-0893764.	
XX			
PA		(BOEH) BOEHRINGER INGELHEIM INT GMBH.	
XX			
Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;			
PI			
XX			
DR		WPI; 1999-120887/10.	
XX			
DR		P-PSDB; AAY00770.	
XX			
PT		New nucleic acid encoding tag7 - used to inhibit tumour growth and	
XX		induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and	
PT		leukaemia	
XX			
PS		Claim 3; Fig 1; 138pp; English.	
XX			
CC		This sequence encodes the murine tag7 of the invention. Cells containing	
XX		the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used	
CC		to produce and purify antibodies; to inhibit growth of mammalian tumours,	
XX		especially for treating carcinoma (of liver, ovary, breast, cervix, lung,	
CC		prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head	
XX		and neck, squamous cell carcinoma or leiocarcinoma), sarcoma (Kaposi's,	
CC		osteosarcoma or fibrosarcoma), melanoma or leukaemia; and as a molecular	
XX		weight marker. The tag7 polypeptide inhibits tumour growth and induces	
CC		apoptosis. The tag7 coding sequences are also useful as probes for gene	
XX		mapping and detection of tag7 gene expression, and as primers. Antibodies	
CC		against tag7 are used as reagents for detecting tag7; as an antagonist of	
CC		tag7; for isolating tag7 and therapeutically to inhibit or delay tumour	
XX		metastasis.	
SO		Sequence 549 BP; 117 A; 161 C; 151 G; 120 T; 0 other;	
Query Match	44.1%;	Score 316.6; DB 20; Length 549;	
Best Local Similarity	77.2%;	Pred No. 8; 8e-71;	

XX	RESULT 11	
XX	AAA80662	
ID	AAA80662 standard; cDNA: 285 BP.	
XX		
XX	AAA80662:	
XX		
XX	21-NOV-2000 (first entry)	
XX		
XX	Human secreted protein gene #8 related gene HBMTB79R SEQ ID #115.	
XX		
KM	Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;	
KM	antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;	
KM	antigenic; vulnery; antiviral; antibacterial; antifungal;	
KM	immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;	
KM	multiple sclerosis; inflammatory disorder; inflammatory bowel disease;	
KM	Crohn's disease; nephritis; hyperproliferative disorder;	
KM	cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;	
KM	melanoma; lymphoma; wound healing; human; ss.	
XX		
OS	Homo sapiens.	
XX		
XX	MO200029435-A1.	
XX		
PD	25-MAY-2000.	
XX		
XX	27-OCT-1999; 99WO-US25031.	
XX		
PR	28-OCT-1998; 98US-0105971.	
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
PI	Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y,	
PI	Greene JM;	
XX		
DR	WPI: 2000-387742/33.	
XX		
PT	Isolated nucleic acid molecules encoding human secreted proteins are	
PT	used for the prevention, amelioration and treatment of autoimmune,	
PT	inflammatory, hyperproliferative and cardiovascular disorders, cancer,	
XX	wounds, and infectious diseases -	
XX		
PS	Disclosure; Page 757; 803pp; English.	
XX		
CC	The present invention relates to 12 secreted human proteins and the	
CC	nucleotide sequences encoding them. The polynucleotide sequences given	
CC	in AAA80606-880623 encode the 12 secreted protein sequences given in	
CC	AA825576-825593. The human secreted proteins have various activities	
CC	dependent on the tissues in which they are expressed. Examples of the	
CC	activities of the proteins include: immunosuppressant;	
CC	anti-inflammatory; antiarthritic; antirheumatic; dermatological;	
CC	antiproliferative; antiarteriosclerotic; anticancer; vulnery;	
CC	antiviral; antibacterial; and antifungal activity. The proteins,	
CC	polypeptides, agonists and antagonists may be used to treat prevent	
CC	and/or diagnose various disease, disorders and conditions examples of	
CC	which include: immune disorders e.g. Addison's disease, rheumatoid	
CC	arthritis, dermatitis, and multiple sclerosis; inflammatory disorders	
CC	e.g. inflammatory bowel disease, Crohn's disease and nephritis;	
CC	hyperproliferative disorders such as paraproteinemias and purpura;	
CC	cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;	
CC	cancer e.g. melanoma and lymphoma. The proteins and polynucleotide	
CC	sequences may also be used in wound healing and the treatment of	
CC	infectious diseases. The human secreted protein gene #8 and protein	
CC	sequences are represented in sequences AAA80613 and AA825583. Sequences	
XX	AAA80662-880663 represent genes related to the secreted protein gene#8.	
XX		
SO	Sequence 285 BP; 71 A; 99 C; 62 G; 47 T; 6 other;	
	Query Match 32.2%; Score 231; DB 21; Length 285;	
	Best Local Similarity 99.2%; Pred. No. 3.9e-49;	
	Matches 242; Conservative 0; Mismatches 1; Indels 1; Gaps 1;	
Y	475 ACCCGAGCCATCCGGCGACGCCAGGGGTCTACTGCGCTGCGGTGAGGAGCCCT 534	

Db	1	ACCCAGGCCATCCGGGAGCCCAAGGGTCTACTGCGCTCGGGTGGCTCAGGGAG-CTT	59
QY	535	GAGGTCCAACTATGTGCTCAAAAGGACACCGGGATGTGACGGTACACTCTCCAGGCAA	594
Db	60	GAGGTCCAACTATGTGCTCAAAAGGACACCGGGATGTGACGGTACACTCTCCAGGCAA	119
QY	595	CCAGCTTACCACTCATCCAGAAATTGGCCACACTACCGTCCCGCTGAGGCCCTGCTGA	654
Db	120	CCAGCTTACCACTCATCCAGAAATTGGCCACACTACCGTCCCGCTGAGGCCCTGCTGA	179
QY	655	TCCGACACCCATTCCTCCCTCCCATGGCCAAAACCCACATGCTCTCTCTCCAAATAA	714
Db	180	TCCGACACCCATTCCTCCCTCCCATGGCCAAAACCCACATGCTCTCTCTCCAAATAA	239
QY	715	GATG 718	
Db	240	GATG 243	
RESULT 12			
ID	AAA80663		
AC	AAA80663	standard; cDNA; 457 BP.	
XX	AAA80663;		
DT	21-NOV-2000	(first entry)	
DE	Human secreted protein gene #8 related gene HODDP40R SEQ ID #116.		
XX	Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;		
KM	antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;		
KM	antitumor; vulnery; antiviral; antibacterial; antifungal;		
KM	immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;		
KM	multiple sclerosis; inflammatory disorder; inflammatory bowel disease;		
KM	Crohn's disease; nephritis; hyperproliferative disorder;		
KM	cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;		
KM	melanoma; lymphoma; wound healing; human; ss.		
OS	Homo sapiens.		
PN	WO200029435-A1.		
XX	25-MAY-2000.		
XX	27-OCT-1999;	99WO-US25031.	
XX	28-OCT-1998;	98US-0105971.	
XX	(HDMA-) HUMAN GENOME SCI INC.		
XX	NI J, Ruben SM, Olsen HS, Young PE, Kenny JT, Moore PA, Wei Y;		
XX	Greene JM;		
XX	WPI; 2000-387742/33.		
XX	Isolated nucleic acid molecules encoding human secreted proteins are		
XX	used for the prevention, amelioration and treatment of autoimmune,		
XX	inflammatory, hyperproliferative and cardiovascular disorders, cancer,		
XX	wounds, and infectious diseases -		
XX	Disclosure; Page 759-760; 803pp; English.		
XX	The present invention relates to 12 secreted human proteins and the		
XX	nucleotide sequences encoding them. The polynucleotide sequences given		
XX	in AAA80663-A80623 encode the 12 secreted protein sequences given in		
XX	AA825576-B25593. The human secreted proteins have various activities		
XX	dependent on the tissues in which they are expressed. Examples of the		
XX	activities of the proteins include: immunosuppressant;		
XX	anti-inflammatory; antiarthritic; antirheumatic, dermatological;		
XX	antiproliferative; antiarteriosclerotic; anticancer; vulnery;		
XX	antiviral; antibacterial; and antifungal activity. The proteins,		
XX	polypeptides, agonists and antagonists may be used to treat prevent		
XX	and/or diagnose various disease, disorders and conditions examples of		

Query Match	30.6%	Score 220;	DB 21;	Length 457;
Best Local Similarity	84.5%	Pred. NO. 2.7e+46;		
Matches 283; Conservative	0;	Mismatches 47;	Indels 5;	Gaps 4

RESULT 13
ABL25883
ID ABL25883 standard; DNA; 558 BP.

PI New Isolated nucleic acid detection reagent for detecting 1000 or more

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB01840-AB16175) and the encoded proteins
CC (AB57737-AB872012).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

50 Sequence 558 BP; 108 A; 199 C; 159 G; 92 T; 0 other;

Query Match	20.5%	Score 147.4	DB 23	Length 558
Best Local Similarly	57.3%	Pred. No. 7.4e-28		
Matches 287	Conservative	0	Mismatches 211	Indels 3
				Gaps 1

QY	129	ACCGGCGCTGCTGCAGGCCCATATATGCGCCCGGAAACAGATGAGAGGCCCTTGCCATCAGAGT	188
Db	50	ACATGGCCACGAGGGGCTGTATGTCCTTCCAAAGGCGAGTGGGGGTGTCGCGGCGCCCAAT	109
QY	189	GCGCCACGACCTTAGCGCTGCGCTTAACTGATGTTGGGGTATGTCACACCGCGGGCAGCA	248
Db	110	GGACCGTAGGCTTGCGCACTACTCAGCTACGCTACGCTATCATCCACACGCGCGGCTCT	169
QY	249	GCTGCAACACCCCGCCCTTGCTGTCAGCAGCAGAGCCCGGAATGTGCACACTACACATCA	308
Db	170	ACTGCGAGACCCGTCGCGCAGTGCAGACCCGCTGTCGAGACGTCACAAATACCACTTGG	229
QY	309	AGACACTGGGCTGGTGGCAGCTGGGGCTACAACTTCTCTGAATGGAGAGACGGGCTCGAT	368
Db	230	ACTCTCTGGGGCTGGCCGAGCATGCGGCTACAACTTCTGATCGGGGGAGAGGCAAGCTGT	289
QY	369	ACGAGGGGCGGTGGGAACCTTACAGGGGTGCCACTAGGTCACCTTATGTGAACCCCATGT	428
Db	290	ACGAGGGGCGGTGGGAACACATGGGCGCCACCGCGCG--CCGAGTGGAAACCCCTACAA	346
QY	429	CCATGGCATCAGCTTATGATGGGCACTACATGATGGGGTGGCCACACCCCGAGCCATCC	488
Db	347	GCACTGGCATCAGCTTCTCTGGCACTTACAACTGGGAGCACCTCTGAGCCGACAAATGATCT	406
QY	489	GGGCAGCCACGAGGCTCTACTGGCCCTGGGAGTGTGGCTCAGGAGGCCCTGAGGTCCAACTATG	548
Db	407	CCGCGCGCCACGACAGCTGTCTCAACGACGCGCCTCAACCTGGCCAGCTCAGTCTCCGGCTACA	466
QY	549	TGCTCAAAAGACACCGGGATGTGCAGCGTACACTCTCTCCAGGCAACCAACCATCTACACC	608
Db	467	TTCCTGTACGGTCAATGCGCAGGTACGGCCACGAAATGCCCCCGCACCCACATCTGGAAAG	526
QY	609	TCAATCCAGATTTGGCCACACT 629	
Db	527	AGATCCGCGGCTGTCCCACT 547	

RESULT 14
ABL25882/C
ID ABL25882 standard; DNA; 2558 BP.

DT	26-MAR-2002 (first entry)
XX	
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 29119.
XX	
KX	Drosophila: developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ds.

XXXXXX


```
Db 3726 ACTCCCTGGGCGTGGCCCGGACATCGGCTACACTTCTGATCGGCGGAGACGGCAACGTGT 3667
OY 369 ACGAGGCGCGTGGCTGGAACTTCACGGGTGCCCACTATGGAACCCCATGT 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3666 ACGAGGCGCGTGGCTGGAACTTCACGGGTGCCCACTATGGAACCCCATGT 3610
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 429 CCATTGGCATCAGCTTCATGGGCACTACATGATGGGTGCCACACACCCAGGCCATCC 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3609 GCATGGGCAATCAGCTTCATGGGCACTACATGATGGGTGCCACACACCCAGGCCATCC 3550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 489 GGGCAGCCCGAGGCTCTACTGGGCTGTGTGCTCAGGGAGGCCCTGAGGTCCAACTATG 548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3549 CCGCGCGCCAGAGAGCTGCTCAACGAGAGCCGCTCAACCGTGGCCAGCTCAGCTCGGCTACA 3490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 549 TGGCTAAAGGACACCGGGATGTGACAGCGTACACTCTCTCCAGGCAACCAAGCTTACCACC 608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3489 TCCTGTACGCTCATCGGCAAGTCAAGCGCCACCGAATGCCCGGACACCATCTGTGGAAG 3430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 609 TCATCCAGAAATGGGCAACT 629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3429 AGATCCGCGGCTGTGCCACT 3409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: November 12, 2002, 00:50:55
Job time : 312 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 12, 2002, 02:29:10 ; Search time 309 seconds
(without alignments)
1392.012 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051

Sequence: 1 MLAWALPSLRIGAAQETE.....LSPGNQLYHLIONMPPHYRSP 191

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL-frame+ p2n.model -DEV-xlp
-O/cgnt2.1/pspro.spool/us09462625/runat_07112002.101525.24830/app.query.fasta.1.327
-DB_N_Geneseq_101002 -QFMT-fastap -SUFIX-rng -MINMATCH=0.1 -LOOPEL=0
-LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=ptio -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us09462625 @CGN.1.1.281 @runat_07112002.101525.24830 -NCPUT=6 -ICPUT=3
-NO_XLPPY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120
-MARKTIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELExt=7

Database :
N_Geneseq_101002: *
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1051	100.0	697	21 AAC58104	Human PRO1269 uncl
2	1051	100.0	697	21 AAC37082	Human PRO1269 (UNO
3	1051	100.0	697	22 AAF54356	DNA encoding prote
4	1051	100.0	718	20 AAX21620	Human tag7 clone c
5	1051	100.0	726	21 AAA80613	Human Htag7 secret
6	1051	100.0	749	21 AAA51719	Chondrosarcoma pep
7	717	68.2	688	18 AAT78509	Bovine granulocyte
8	692	65.8	549	20 AAT21819	Mouse tag7 clone c
9	648.5	61.7	677	18 AAT78510	Murine granulocyte
10	573	54.5	380	21 AAT42549	Human 5' EST Isola
11	414.5	39.4	555	23 ABL25881	Drosophila melanog
12	414.5	39.4	2555	23 ABL25880	Drosophila melanog
13	413	39.3	726	24 ABA90336	Human polynucleoti
14	413	39.3	1110	22 ABA04006	Human full length
15	413	39.3	1110	24 ABA90337	Human polynucleoti
16	413	39.3	1128	23 ABA04004	Human full length
17	405.5	38.6	558	23 ABL25883	Drosophila melanog
18	405.5	38.6	639	23 ABL04835	Drosophila melanog
19	405.5	38.6	2558	23 ABL25882	Drosophila melanog
20	405.5	38.6	2639	23 ABL04834	Drosophila melanog
21	405.5	38.6	4320	23 ABL25876	Drosophila melanog
22	395.5	37.6	1876	21 AAA51718	Wound healing tiss
23	379	36.1	612	23 ABL26901	Drosophila melanog
24	371.5	35.3	457	21 AAA80663	Human secreted pro
25	348.5	33.2	1089	21 AAA15921	Human protein clon
26	348.5	33.2	1256	21 AAA15921	Human protein clon
27	348.5	33.2	1731	22 AAT67206	Nucleotide sequenc
28	348.5	33.2	1905	22 AAT67205	Nucleotide sequenc
29	348	33.1	753	19 AAV19149	Gene encoding the
30	347.5	33.1	591	19 AAV19148	Gene encoding the
31	347	33.0	1256	23 ABL14655	Drosophila melanog
32	341	32.4	1876	22 AAD02743	Murine peptidoglyc
33	339	32.3	1107	22 AAD04007	Human full length
34	339	32.3	1125	22 AAD04005	Human full length
35	338	32.2	2609	23 ABL15556	Drosophila melanog
36	338	32.2	2641	23 ABL15528	Drosophila melanog
37	331.5	31.5	9721	23 ABL07586	Drosophila melanog
38	330	31.4	561	23 ABL08767	Drosophila melanog
39	330	31.4	2561	23 ABL08766	Drosophila melanog
40	329.5	31.4	573	23 ABL15529	Drosophila melanog
41	328.5	31.3	522	19 AAV19147	Gene encoding pept
42	327	31.1	1014	23 ABL07655	Drosophila melanog
43	317	30.2	970	23 ABL25789	Drosophila melanog
44	316.5	30.1	1182	21 AAA51717	Keratinocyte pepti
45	316.5	30.1	1191	21 AAT65250	Human secreted pro

ALIGNMENTS

RESULT 1
AAC58104
ID AAC58104 standard; cDNA; 697 BP.
AAC58104;
25-JAN-2001 (first entry)
Human PRO1269 nucleotide sequence SEQ ID NO:6.
Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
identification; tumorigenesis; anticancer; detection; ss.
Homo sapiens.
XX
XX WO200053750-A1.
XX
XX
XX
PD 14-SEP-2000.

XX 02-DEC-1999; 99WO-US28551.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WT;
XX WPI; 2000-594320/56.
DR P-PSDB; AAB24022.
XX
XX Antibodies specific for pro polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of pro
PT polypeptide activity or expression -
XX
PS Claim 50; Fig 3; 226pp; English.
XX
CC The present invention describes an antibody that binds to a human
CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO334; PRO1927; PRO3567; PRO1293; PRO1303; PRO4354; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC anticancer activity and can be used to diagnose tumors in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumors. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridizes to genes encoding (I), can be used to inhibit tumour growth,
CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human pro
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC pro polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
XX Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 5.17e-100 Length: 697
Score: 1051.00 Matches: 191
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-462-625-4 (1-191) x AAC58104 (1-697)
QY 1 MetleuAlaTATPAlaLeuProSerleuArgleuGlyAlaAlaGlnIurhgu 20
DB 41 ATGCTGCTGGCTGGGCTCTCCAGCTCTCTGCTGAGGCGGCTAGAGACAGAA 100
QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu 40
DB 101 GACCGGCGCTGCTGACGCCCATAGTGCCTCCGAGACGAGTGAAGCCCTGCGCATCAGAG 160
QY 41 CysAlaGlnHisLeuSerleuProleuArgTrpValValSerHisThrAlaGlySer 60
DB 161 TGGCGCCAGCAGCTAGGCTGCTTACGCTATGTGTGGTATCCACACAGCGGCGCAGC 220
QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisTrpHisMet 80
DB 221 AGCTGCAACACCCCGCTCGTCCAGCAGCGAGCGGAGATGTGACGACATACCATG 280
QY 81 LysThrLeuGlyTrpCysAspValGlyTrpAsnPhaLeuIleGlyGluAspGlyLeuVal 100
DB 281 AAGACACTGGGCTGTGGAGCTGGGCTACACTCTCTATGTGAGAAACAGCGGCTCGTA 340
QY 101 TyrGluGlyArgGlyTrpAsnPhaThrGlyAlaHisSerGlyHisLeuTrpAsnProMet 120

DB 341 TACGAGGCGCGCTGGCTGGAATTCACGGGCTGCCACCTCAGGTCACCTATGAAACCCCATG 400
QY 121 SerIleGlyIleSerPheMetGlyAsnTrpMetAspArgValProThrProGlnAlaIle 140
DB 401 TCCATTGGCATGCTCTCATGGCACTACATGATCGGGTGGCCACACCCAGGCCATC 460
QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTrp 160
DB 461 CGGCGACGCCAGGCTCTATGCGCTGCGGTGTGGCTCAGGAGCCCTGAGGTCCACTAT 520
QY 161 ValLeuLysGlyHisArgAspValGlnArgTrpLeuSerProGlyAsnGlnLeuTrpHis 180
DB 521 GTGCTCAAGGACACCGGCGATGCGAGCTTACACTCTCTCAGGCAACACGCTCTACAC 580
QY 181 LeuIleGlnAsnTrpProHisTrpArgSerPro 191
DB 581 CTCATCCAGAAATTGGCCACACTACCGCTCCCC 613
RESULT 2
ID AAA37082 standard; cDNA; 697 BP.
XX
AC AAA37082;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (U0639) cDNA sequence SEQ ID NO:215.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW Transmembrane; secretion; Immunoadhesion; pharmaceutical; screening;
KW ss.
XX
OS Homo sapiens.
XX
PN W0200012708-A2.
XX
PD 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
PF 01-SEP-1998; 98US-0098716.
XX 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.

QY 101 TyrgluglyAArglyTTPAsnPhenrGlyAlaHisSerGlyHisLeuTTPAsnPromet 120
 |||||
 Db 341 TACAGGGCCCTGGTGGAACTTCACGGGTGCCCTCAGGCTACTTATGGAACCCCATG 400
 QY 121 SerilegIyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 |||||
 Db 401 TCCATTGGCATCAGCTTCATGGGCAACTACATGATCGGGTCCCAACCCAGGCATAC 460
 QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
 |||||
 Db 461 CGGGAGCCCAAGGCTCTACGGCTGCGGTGGCTCAGGGGCCCTGAGGTCCAACTAT 520
 QY 161 ValLeuIySGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 |||||
 Db 521 GTGCTCAAGAGACACCGGAGTGTGACGGTACACTCTCTCCAGGCAACAGCTTACCA 580
 QY 181 LeuIleGlnAsnTTPProHisTyrArgSerPro 191
 |||||
 Db 581 CTCATCCAGAAATTGGCCACACTACCGCTCCCCC 613

RESULT 3

AAF54356
 ID AAF54356 standard; DNA; 697 BP.

XX AAF54356;

DT 02-APR-2001 (first entry)

DE DNA encoding protein of the invention #61.

XX Secreted; transmembrane; gene therapy; ss.

KW Unidentified.

OS WO200078961-A1.

XX 28-DEC-2000.

PF 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

FR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US30095.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

XX (GETH) GENENTECH INC.

PI Baker KP, Bolstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

PI Pan J, Peoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

PI Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-071395/08.

PT Secreted and transmembrane proteins and nucleic acids designated PRO,

PT useful as hybridization probes, in chromosome and gene mapping and gene

PT therapy -

XX Alignment Scores:
 Seq Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
 Pred. No.: 5.17e-100
 Score: 1051.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 22
 Gaps: 0
 US-09-462-625-4 (1-191) x AAF54356 (1-697)
 QY 1 MetLeuLeuAlaTTPAlaLeuProSerLeuLeuArgGlyValAlaGlnGluThrGlu 20
 |||||
 Db 41 ATGCTGCTTGGCTGGGCTCTCCCAAGCTCTTCGACTGGAGGGCTTACAGACAGAA 100
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTTPlysAlaLeuAlaSerGlu 40
 |||||
 Db 101 GACCGGCTGCTGCAGCCCATAGTGGCCGGAACGAGTGAAGGCCCTGGCATCAGAG 160
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyrValValAlaSerHisThrAlaGlySer 60
 |||||
 Db 161 TCGGCCCAAGCACCTGAGCCCTTACGCTATGTGTGTATCCACACAGCGGGCAGAG 220
 QY 61 SerCysAsnThrProAlaSerCysGlnGlnAlaArgAsnValGlnHisTyrHisMet 80
 |||||
 Db 221 AGCTGCACACCCCGCCCTGTCGACAGCAGAGCCCGGAATGTGCACACTACACATG 280
 QY 81 LysThrLeuGlyTTPCysAspValGlyTyrAsnPhenrGlyAlaHisSerGlyHisLeuTTPAsnPromet 100
 |||||
 Db 281 AAGACACTGGGCTGGTGGCAGCTGGGCTTCACTTCCATTTGGAAGACCGGCTCGTA 340
 QY 101 TyrgluglyAArglyTTPAsnPhenrGlyAlaHisSerGlyHisLeuTTPAsnPromet 120
 |||||
 Db 341 TACAGGGCCCTGGTGGAACTTCACGGGTGCCCTCAGGCTACTTATGGAACCCCATG 400
 QY 121 SerilegIyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 |||||
 Db 401 TCCATTGGCATCAGCTTCATGGGCAACTACATGATCGGGTCCCAACCCAGGCATAC 460
 QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
 |||||
 Db 461 CGGGAGCCCAAGGCTCTACGGCTGCGGTGGCTCAGGGGCCCTGAGGTCCAACTAT 520
 QY 161 ValLeuIySGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 |||||
 Db 521 GTGCTCAAGAGACACCGGAGTGTGACGGTACACTCTCTCCAGGCAACAGCTTACCA 580
 QY 181 LeuIleGlnAsnTTPProHisTyrArgSerPro 191
 |||||
 Db 581 CTCATCCAGAAATTGGCCACACTACCGCTCCCCC 613

RESULT 4

AAAX21820

ID AAAX21820 standard; cDNA to mRNA; 718 BP.

XX AAAX21820;

DT 18-MAY-1999 (first entry)

DE Human tag7 clone coding sequence.

XX Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;

XX melanoma; leukaemia; apoptosis inducer; human; ss.

XX Homo sapiens.

XX WO9902686-A1.

XX 21-JAN-1999.

PD 10-JUL-1998; 98WO-BP04287.

XX

XX

XX

XX 11-JUL-1997; 97US-0893764.
 PR (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 PA Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
 XX WPI: 1999-120887/10.
 DR P-PSDB: AAY00771.
 XX
 PT New nucleic acid encoding tag7 - used to inhibit tumour growth and
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
 PT leukaemia
 PS
 PS Claim 11: Page 126-127; 138pp; English.
 XX
 CC This sequence encodes the human tag7 of the invention. Cells containing
 CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
 CC to produce and purify antibodies; to inhibit growth of mammalian tumours,
 CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
 CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
 CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
 CC osteo- or fibro-sarcoma), melanoma or leukaemia; and as a molecular
 CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
 CC apoptosis. The tag7 coding sequences are also useful as probes for gene
 CC mapping and detection of tag7 gene expression, and as primers. Antibodies
 CC against tag7 are used as reagents for detecting tag7; as an antagonist of
 CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
 CC metastasis.
 CC
 XX
 XX Sequence 718 BP; 142 A; 251 C; 194 G; 131 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 5.39e-100 Length: 718
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 20 Gaps: 0
 US-09-462-625-4 (1-191) * AAX21820 (1-718)
 QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGlyAlaAlaGlnGluThrGln 20
 Db 68 ATGCTGCTTGCGCTGGCGCTCTCCAGCCCTCTTCAGTGGAGCGGCTAGAGACAGAA 127
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGln 40
 Db 128 GACCCGGCGCTGCTGCAGCCCTCATAGTCCCGAAGAGAGCCCTGCGATCAGAG 187
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTrpValValSerHisAspThrAlaGlySer 60
 Db 188 TGGCGCCAGCACCTAGAGCTGCGCTTACGCTATGTTGTTGTCACACGCGCGGCACG 247
 QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHisMet 80
 Db 248 AGCTGCAACACCCCGGCTCTGTCGACAGCAGCGCCGGAATGTGACAGACTACACATG 307
 QY 81 LysThrLeuGlyTrpCysAspValGlyTrpAsnPhenLeuIleGlyGlnAspGlyLeuVal 100
 Db 308 AAGACACTGGCGCTGCTGCAGCGTGGGCTACACTTCCTGATGGAAGACGGGCTCGTA 367
 QY 101 TyrGlnGlyArgGlyTrpAsnPhenThrGlyAlaHisSerGlyHisLeuThrAsnProMet 120
 Db 368 TACGAGGCGCGGCTGGAATCTACAGGGTGCCTCAGGCTACTATGGAACCCCATG 427
 QY 121 SerIleGlyIleSerPheMetGlyAsnTrpMetAspArgValProThrProGlnAlaIle 140
 Db 428 TCCATTGGCATATAGCTTCAATGCGCAACTACATGATCGGGTCCGCCACCCAGCGCATC 487
 QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTrp 160
 Db 488 CGGCGAGCCCGAGGCTCTACTGCTGCGGTGTGGCTCAGAGGACCCCTGAGGTCCACTAT 547

QY 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTrpHis 180
 Db 548 GTGCTCAAGAGACACCGGAGTGTGCGAGCTACACTCTCTCAGGACACGACTTACAC 607
 QY 181 LeuIleGlnAsnTrpProHisTrpArgSerPro 191
 Db 608 CTCATCCAGAAATTGGCGCACATCACGCGCTCCGCC 640
 RESULT 5
 ID AAA80613 standard; cDNA; 726 BP.
 AC AAA80613;
 XX 21-NOV-2000 (first entry)
 DE Human Htag7 secreted protein gene #8.
 XX
 KM Secreted protein; immunosuppressant; anti-inflammatory; antirheumatic;
 KM antirheumatic; dermatological; antiproliferative; antiartherosclerotic;
 KM anticancer; vulnary; antiviral; antibacterial; antifungal;
 KM immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KM multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KM Crohn's disease; nephritis; hyperproliferative disorder;
 KM cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KM melanoma; lymphoma; wound healing; human; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200029435-A1.
 XX
 PD 25-MAY-2000.
 XX
 PF 27-OCT-1999; 99WO-US25031.
 XX
 PR 28-OCT-1998; 98US-0105971.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 NI J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wel Y;
 PI Greene JM;
 XX
 DR WPI: 2000-387742/33.
 DR P-PSDB: AAB25583.
 XX
 PT Isolated nucleic acid molecules encoding human secreted proteins are
 PT used for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PT wounds, and infectious diseases -
 PS
 PS Claim 1: Figure 34; 803pp; English.
 PS
 CC The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given
 CC in AAA80606-A80623 encode the 12 secreted protein sequences given in
 CC AAB25576-B25593. The human secreted proteins have various activities
 CC dependent on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant;
 CC anti-inflammatory; antirheumatic; antihemmatic; dermatological;
 CC antiproliferative; antiartherosclerotic; anticancer; vulnary;
 CC antiviral; antibacterial; and antifungal activity. The proteins,
 CC polypeptides, agonists and antagonists may be used to treat prevent
 CC and/or diagnose various disease, disorders and conditions examples of
 CC which include: immune disorders e.g. Addison's disease, rheumatoid
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 CC hyperproliferative disorders such as paraneoplasms and purpura;
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 CC sequences may also be used in wound healing and the treatment of
 CC infectious diseases. The human secreted protein gene #8 and protein
 CC sequences are represented in sequences AAA80613 and AAB25583. Sequences

CC AAA0662-A80663 represent genes related to the secreted protein gene#8.
 XX Sequence 726 BP; 161 A; 251 C; 188 G; 126 T; 0 other;
 SQ Alignment Scores:
 Pred. No.: 5,47e-100 Length: 726
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-462-625-4 (1-191) x AAA80613 (1-726)

QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGlnAlaAlaGlnGluThrGlu 20
 DB 47 ATGCTCTTGGCTGGGCTCTCCAGCCTCTTCGACTCGAGCGGCTGAGAGACGAA 106
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpPylsAlaLeuAlaSerGlu 40
 DB 107 GACCCGGCTGCTGCGAGCCCATAGTCCCGGAGAGAGTGAAGGCCCTGGCATCAGAG 166
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyTyrValValSerHisThrAlaGlySer 60
 DB 167 TGGCCAGACAGCTGAGCCCTGACCTTACGCTATGCTGCTATCGCACAGCGCGGCGAGC 226
 QY 61 SerCysAsnThrProAlaSerCysGlnGlnAlaArgAsnValGlnHisThrHisMet 80
 DB 227 AGGTGGAAACCCCGCTGCTGCGACAGACAGCCCGGAATGGCGACACTACCATATG 286
 QY 81 LysThrLeuGlyTrpCysAspValGlyTyrAsnPheLeuIleGlyGluAspGlyLeuVal 100
 DB 287 AAGACACTGGGCTGCTGCGAGCGTGGCTACAACTCTGATTGGAAGAAGCGGCTCGTA 346
 QY 101 TyrGluGlyArgGlyTrpAsnPheThrGlyAlaHisSerGlyHisLeuTrpAsnProMet 120
 DB 347 TACGAGGCGCTGCTGGAACCTTACCGGCTGCCACACTGCTACTATGGAACCCCATG 406
 QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 DB 407 TCCATTGGCAATCACTTCAATGCGCACTACATGATGGGTGCCACACCCAGGCCATC 466
 QY 141 ArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaGlnGlnGlyAlaLeuArgSerAsnTyr 160
 DB 467 CGGGACAGCCAGGCTACAGCGCTGCGGTGGCTCAGAGGAGCCCTGAGSTCCAACTAT 526
 QY 161 ValLeuGlyGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 DB 527 GTGCTCAAAAGGACACCGGAGTGTGACAGCTCTCCAGGCAACCACTCTACCCAC 586
 QY 181 LeuIleGlnAsnTrpProHisTyrArgSerPro 191
 DB 587 CTCATCCAGAAATGGCCACACTACCGCTCCCC 619

RESULT 6
 AAA51719 standard; cDNA; 749 BP.
 ID AAA51719;
 AC AAA51719;
 XX 31-OCT-2000 (first entry)
 DT Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.
 DE Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.
 XX Peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator;
 KW Chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
 KM tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
 KW inhibitor; protein co-ordinate data; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FH CDS 55..645
 FT CDS

FT FT /tag- a
 FT sig-peptide /product- PGRP-like_protein
 FT 55..117
 FT mat-peptide /tag- b
 FT 118..642
 FT /tag- c

WO200039327-A1.
 06-JUL-2000.
 22-DEC-1999; 99WO-US30736.
 23-DEC-1998; 98US-0113809.
 (HUMAN) HUMAN GENOME SCI INC.
 Rosen CA, Ruben SM, Young PE, Olsen HS;
 WPI; 2000-452414/39.
 P-PSDB; AAY96964.

polynucleotide encoding peptidoglycan recognition protein-like protein,
 antibodies specific to it useful for preventing, treating conditions
 e.g. endotoxemic shock and auto-immune disorders and infections in mammai

PS Claim 1; Fig 3; 191pp; English.

XX Novel human peptidoglycan recognition protein-like proteins (PGRP)
 XX expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
 XX or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
 XX proteins are useful for preventing, treating or ameliorating a medical
 XX condition in a mammal (claimed). PGRP is useful in augmenting the immune
 XX system in such areas as immune recognition, antigen presentation and
 XX immune system activation. Antibodies or antagonists directed against
 XX these proteins may be useful in reducing or eliminating disorders
 XX associated with tumour necrosis factor (TNF) and TNF-like cytokines,
 XX such as endotoxemic shock and autoimmune disorders and for treating
 XX infectious diseases including silicosis, sarcoidosis and idiopathic
 XX CC pulmonary fibrosis.

SQ Sequence 749 BP; 165 A; 259 C; 195 G; 130 T; 0 other;
 SQ Alignment Scores:
 Pred. No.: 5,7e-100 Length: 749
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-462-625-4 (1-191) x AAA51719 (1-749)

QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGlnAlaAlaGlnGluThrGlu 20
 DB 70 ATGCTCTTGGCTGGGCTCTCCAGCCTCTTCGACTCGAGCGGCTGAGAGACGAA 129
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpPylsAlaLeuAlaSerGlu 40
 DB 130 GACCCGGCTGCTGCGAGCCCATAGTCCCGGAGAGAGTGAAGGCCCTGGCATCAGAG 189
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyTyrValValSerHisThrAlaGlySer 60
 DB 190 TGGCCAGACAGCTGAGCCCTTACGCTATGCTGATGCGACACAGGCGGAGC 249
 QY 61 SerCysAsnThrProAlaSerCysGlnGlnAlaArgAsnValGlnHisTyrHisMet 80
 DB 250 AGCTGCAACACCCCGCTGCTGCGAGCAGAGCGCCGGAATGTGACACACTACCATAG 309
 QY 81 LysThrLeuGlyTrpCysAspValGlyTyrAsnPheLeuIleGlyGluAspGlyLeuVal 100
 DB 310 AAGACACTGGGCTGCTGCGAGCGTGGCTACAACTCTGATTGGAAGAAGCGGCTCGTA 369

	US-09-462-625-4 (1-191) x AAT78509 (1-688)	Alignment Scores:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
CC	acids corresponding to mature BGP-A. BGP-A and the murine		2,84e-65	717.00	80.75%	70.05%	18
CC	homolog, MGP-A (see AAW3375), exhibit activity against			Matches: 131	Conservative: 20	Mismatches: 30	1
CC	Gram-positive and Gram-negative bacteria, fungi and viruses,						
CC	specifically <i>Staphylococcus aureus</i> and <i>Escherichia coli</i> , <i>Candida</i>						
CC	<i>albicans</i> , <i>Salmonella typhimurium</i> and <i>C. neoformans</i> (claimed).						
CC	They can be used in human or veterinary medicine (particularly to						
CC	treat disorders associated with lipopolysaccharides, e.g. sepsis						
CC	and endotoxaemia) or as preservatives in food products or in water						
CC	supplies (claimed). They can also be applied to crops to reduce						
CC	post-harvest spoilage or expressed in transgenic plants to increase						
CC	their disease resistance. They have low immunogenicity.						
XX							
XX	Sequence 688 BP; 126 A; 234 C; 201 G; 127 T; 0 other:						
US-09-462-625-4 (1-191) x AAT78509 (1-688)							
OY	3 LeuAlaTrpAlaLeuProSerLeuLeuArgLeuGlnAlaAlaGlnGlnTrpGluAspPro						
DB	50 CTCGGCTGGGCTCCCTCCGCTCTCGGCTCGGGCGGCTCAAGC-----						
OY	23 AlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGluCysAla						
DB	98 -----TCCGGCAGCATCTGTTCCCGGGAAAGGGGGCGCCCGGATCCAAATGACG						
OY	43 GlnHisLeuSerLeuProLeuArgTrpValValValSerHisThrAlaGlySerSerCys						
DB	152 CAGAGGCTAAACACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG						
OY	63 AsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHisMetLysThr						
DB	212 AACACTCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG						
OY	83 LeuGlyTrpCysAspValGlyTrpAsnPheLeuIleGlyGluAspGlyLeuValTrpGlu						
DB	272 CGGGCTGGTGGCAGCGTGGCTACAAATTTCTGATCGAGAAATGGGCTGCTATAGAG						
OY	103 GlyAlaGlyTrpAsnPheThrGlyAlaHisSerLeuHisLeuTrpAsnPrometSerIle						
DB	332 GCGCGGGGCTGGAACACCTTAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG						
OY	123 GlyIleSerPheMetGlyAsnTrpMetAspArgValProThrProGlnAlaIleArgAla						
DB	392 GGCAATCTCTTCATGGGGAACTACATGATGGGTGGCCCCCGGCTCTGCTCAGGGCG						
OY	143 AlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTrpValLeu						
DB	452 GCCAGAGTCTGCTGGCTTGTGGGCGACGCTGGGGATACCTGATCTCACTACGAGAATC						
OY	163 LysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTrpHisLeuIle						
DB	512 AAAAGGACACCGCGATGTGCACAAACGCTCTCTCAGGGGACGAGCTCTATAAATATC						
OY	183 GlnAsnTrpProHisTrpArg 189						
DB	572 CAGCAGTGGCGGCACTACCGC 592						
RESULT 8							
ID	AAx21819						
AC	AAx21819 standard; CDNA; 549 BP.						
XX	AAx21819;						
XX	18-MAY-1999 (first entry)						
DE	Mouse tag7 clone coding sequence.						

Oy		101	TyrIlnclnglyAagGylTTPaanPheThrGlyAlaHlsserGlyHlsLeutrPaanPromet	120
Dd		277	TATGAGGCCCGACGGCTGGAACTCAATCAAGGGTGACCACACAGGCCCCCATCTGGATTCCCATC	336
Oy		121	SerlleeglylIserPhemetGlyAsnTrYmetAsparGylProthrProgInalaIle	140
Dd		337	TCFATGTGCATCACCTTCATGAGGGAACTTCATGATGACCGGGTACCCGAAGCGGGCCCTC	396
Oy		141	ArgAlaAlaGlnGlyLeuLeuAlaCygGlyValAlaIlnGlyAlaLeuArgSerAsnTr	160
Dd		397	CGTCTGCCCCCTAAATCTTCTGGAAATGGGGGTGTCTCGGGCTTCAGATCCACTAT	456
Oy		161	ValleuysgVhlSAtGAspValGlArgThlrLeuSerProGlyAsnGlnLeuTryHis	180
Dd		457	GAACTCAAAGACACCGGGATGTCTCAAGAACACTCTCTCCAGGTGACCAACTATACAG	516
Oy		181	LeuileglnAsnTrpTrohisTyArg	189
Dd		517	GTCATCCAAAGCTGGGAACACTACCGA	543
RESULT 9				
AAT78510		ID	AAT78510 standard; cDNA; 677 BP.	
XX		AC	AAT78510;	
XX		DT	18-FEB-1998 (first entry)	
XX		DE	Murine granulocyte peptide A precursor cDNA.	
XX		KW	Antimicrobial peptide; antibiotic; antibacterial; antifungal;	
KW		KM	fungicide; antiprotozoas; protozoacide; antiviral; virucide;	
KW		KN	murine granulocyte peptide A; MGP-A; preservative; sepsis;	
XX		OS	endotoxaemia; mouse; ss.	
XX		MS	Mus musculus.	
XX		FH		
FT		CDS	Location/Qualifiers 36..581	
FT		mat_peptide	/tag= a /trans_except= (pos:561..563, aa:Trp) 543..578 /tag= b /product= MGP-A	
FT				
PN			WO9729765-A1.	
PD			21-AUG-1997.	
XX		PF	13-FEB-1997; 97WO-USO2218.	
XX		PR	16-FEB-1996; 96US-0011834.	
PPA			(REGC) UNIV CALIFORNIA.	
P1			Seisted ME;	
DR			WPI, 1997-424753/39.	
DR			P-PsDB; AAM23723.	
PT				
PT			Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -	
PT			useful therapeutically, as preservatives for food, in water	
PS			treatment and in agriculture	
PS			Claim 14; Fig 5; 56pp; English.	
CC			This cDNA sequence encodes the precursor (see AAM23723) of a novel,	
CC			claimed antimicrobial peptide from bovine neutrophils, designated	
CC			murine granulocyte peptide A or MGP-A (see AAM23725). It was	
CC			isolated from murine bone marrow cDNA using primers based on	
CC			bovine BGP-A cDNA (see AAT78509). The encoded MGP-A precursor	
CC			comprises signal peptide and propeptide sequences followed by 12	
CC			c-terminal amino acids corresponding to mature MGP-A. MGP-A and	

QY	61	SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAspValGlnHisTyrHisMet	80
Db	228	AGCTGCAACACCCSCCGCTCGTGCACGACGACGAGCCCCGGAAATGTGCAGCACTACCAATG	287
QY	81	LysThrLeuGlyTyrCysAspValGlyTyrAspPheLeu-IleGlyLysAspGlyLeuA	100
Db	288	AAGACACACGCGGTGGTGGCAGCTGGGCTCAACACTKCTTGATTTGAGAGACGCGCTCGT	347
QY	100	ITyGluGlyAspGlyTyrAspPheThrGly	110
Db	348	ATACAGAGGCGCGTGGMTGGAACTCAACGCGGT	378
RESULT 11			
ID	ABL25881	standard; DNA; 555 BP.	
XX	ABL25881;		
XX	ABL25881;		
DT	26-MAR-2002	(first entry)	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 29116.		
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ds.		
XX	Drosophila melanogaster.		
OS	WO200171042-A2.		
PN	27-SEP-2001.		
PD	23-MAR-2001; 2001WO-US09231.		
PE	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX			
PA	(PEKE) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
DR	WPI; 2001-656860/75.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions -		
XX			
PS	Claim 1; SEQ ID NO 29116; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
CC	sequences (ABL01840-ABL16175) and the encoded proteins		
CC	(AB857737-AB872072).		
CC	The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SO	Sequence 555 BP; 120 A; 192 C; 140 G; 103 T; 0 other;		
Alignment Scores:			
Pred. No.:	6.26e-34	Length:	555
Score:	414.50	Matches:	77
Percent Similarity:	66.26%	Conservative:	31
Best Local Similarity:	47.24%	Mismatches:	54
Query Match:	39.44%	Indels:	1
DB:	23	Gaps:	1
US-09-462-625-4 (1-191) x ABL25881 (1-555)			

Db	67	ATCATCTCCAAAGTTCGAGATGGGGGCGCGCTTCCGCCACAGCAAGACCTTCGTGGCCAAAC	126
Qy	48	ProteinArgTyrValValValSerHisThrAlaGlySerSerCysAsnThrProAlaSer	67
Db	127	TACCTGAGCTACGCGCTGATCCACACACCGCTGGAAACTACTGCAGCACCAAGGCGCC	186
Qy	68	CysGlnGlnGlnAlaArgAsnValGlnHisTyrHisMetLeuThrLeuGlyTyrPyrCysAsp	87
Db	187	TGCATCCACACACCTGCAGAACATCCAGAGGCTCTCCACATGTGAGTCTCTGGGCTGGCGCGAT	246
Qy	88	ValGlyTyrAsnPheLeuLeuIleGlyGluAspGlyLeuValTyrGluGlyTyrPasn	107
Db	247	ATCGGCTCACTACTTCTCTGATCCGGGAGACACGGCAGACGTATACGAGGTGCGGCTGGAAC	306
Qy	108	PheThrGlyAlaHisSerGlyHisLeuTyrPasnPrometSerIleGlyIleSerPheMet	127
Db	307	GTTATGGGTGCTCAGCGCCACTAAC--TGAACCTCCAAAGTATATGCGGATCTCTTCCTG	363
Qy	128	GlyAsnTyrMetLeuAspArgTyrValProThrPyrProGlnAlaIleArgAlaAlaGlnGlyLeu	147
Db	364	GCCACTACAAATACCAACACCCCTCACTCTGCTCAGATCACCGGCTGCCAAGGCTCTGCTC	423
Qy	148	AlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyrValLeuGlySerGlyHisArgAsp	167
Db	424	TCCGATGCGGCTAGTCGCGGCCAGATCGTTTCCGGATATCATCTCTGTACGACATCCGAC	483
Qy	168	ValGlnArgTyrThrLeuSerProGlyAsnGlnLeuTyrHisLeuIleGlnAsnTyrProHis	187
Db	484	GTCGGCTCCACCGAGTCCCGGACCAACATCTGGAAGAGATCCGCACTGTGTCCAAC	543
Qy	188	TyrArgSer 190	
Db	544	TGGAAGGCC 552	
RESULT 12			
ID	ABL25880/C		
	ABL25880 standard; DNA; 2555 BP.		
XX	ABL25880;		
XX	AC		
XX	DT		
XX	26-MAR-2002 (first entry)		
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 29113.		
XX	XX		
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ds.		
XX	XX		
OS	Drosophila melanogaster.		
XX	XX		
PN	WO200171042-A2.		
PD	XX		
XX	27-SEP-2001.		
PF	23-MAR-2001; 2001WO-US09231.		
XX	XX		
PR	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX	XX		
PA	(PEKE) PE CORP NY.		
XX	XX		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX	XX		
DR	WPI; 2001-656860/75.		
XX	XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
XX	interactions -		
PS	Claim 1; SEQ ID NO 29113; 21pp + Sequence Listing; English.		
XX	XX		
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		

QY 20 GLUASPRoAlAcYsCySeSerProIleValProArgAsnGluTrpIlyAlaLeuAlaSer 39
 DB 217 AGCCTGAAGAAGGCTGCGCCGCGCTTCCACGCGTCTGTGTGGGAGCCAGGAGAC 276
 QY 40 GLUCysAlaGlnHisLeuSerLeuProLeuArgTyrValValSerHisThrAlaGly 59
 DB 277 CACTGTCCAGG---ATGACTCTCCAGGAGATGATGCGCATTAATCCACACTCCGGG 333
 QY 60 SerSerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisTyrHis 79
 DB 334 AGGACCTGCAACATTCTGTATGATGAGTCCGCTGTGTCGGGACATCCAGCTTTCTAC 393
 QY 80 MetLysThrLeuGlyTyrCysAspValGlyTyrAsnPhenLeuIleGlyGluAspGlyLeu 99
 DB 394 ATAGACAGGCTCAAGCTATGCGACATGTGTTATTAATCTCTGTGTGGGACAGATGGCGCC 453
 QY 100 ValTyrGlnGlyArgGlyTyrAsnPhenThrGlyAlaHisSerGlyHisLeuTyrAsnPro 119
 DB 454 ATTATGAAAGGGGTGGGCGGATGTCCAAAGGCTCC---TCCACCCCTGGGTACGATGAC 510
 QY 120 MetSerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAla 139
 DB 511 ATTGCCCTGGGCAATTAATGCGACCTTCACAGGTATGACCAACCCATGCTGACGCA 570
 QY 140 IleArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaIleGlnGlyAlaLeuArgSerAsn 159
 DB 571 CTAGAGGAGCCCAAGACCTGATCCAGTGTGCGCATCAAGGCTGACTGCCCAAC 630
 QY 160 TyrValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyr 179
 DB 631 TACCTGCGTGGGCGCCACAGTGTGTCGCGCAACCTGTCTCCCTGGGACAGCTTTGTAC 690
 QY 180 HisLeuIleGlnAsnTyrProHisTyrArg 189
 DB 691 AACATCATCAGCACCTGGCGCTCATTTCAA 720
 RESULT 14
 AAD04006
 ID AAD04006 standard; CDNA; 1110 BP.
 AC AAD04006;
 XX
 XX 02-JUL-2001 (first entry)
 DE Human full length granulocyte peptide homolog Zgpai cDNA #2.
 KW Human; granulocyte peptide A; GP-A; Zgpai; cytosolic; antiinflammatory;
 KW vulnereary; dermatological; anti-microbial; gastrointestinal disease;
 KW pulmonary; dental caries; periodontal disease; gene therapy; AIDS;
 KW Acquired Immune Deficiency Syndrome; epithelial wound; cystic fibrosis;
 KW Infection; urinary tract; respiratory; vaginal; lung; skin; cancer;
 KW ovarian; rectal; chromosome 1; ss.
 KW
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1110
 FT /tag- a
 FT /product= "Human full length zgpai protein #2 with
 FT linker sequence"
 FT sig_peptide 1..51
 FT /tag- b
 FT mat_peptide 52..1107
 FT /tag- c
 FT /product= "Mature human full length zgpai protein #2"
 XX
 XX WO200129224-A2.
 XX
 XX 26-APR-2001.
 XX
 XX 20-OCT-2000; 2000WO-US29177.
 PF
 XX

PR 20-OCT-1999; 99US-0160712.
 PR 12-JUL-2000; 2000US-0218070.
 XX
 XX
 PA (ZIMO) ZIMOGENETICS INC.
 PI Conklin DC, Adler DA, Fox BA;
 PI MPI; 2001-290918/30.
 DR P-PSDB: AAE00693.
 XX
 XX New granulocyte peptide homolog, zgpai polypeptide, for research
 PT applications, diagnosis and treatment of cancer, periodontal,
 PT gastrointestinal disease, urinary tract, skin and lung infections
 PS Claim 2: Page 107-109; 114pp; English.
 CC The present sequence is a cDNA encoding human full length granulocyte
 CC peptide (GP-A) homolog, zgpai protein. Zgpai gene is located on human
 CC chromosome 1. Zgpai polypeptides are useful for producing antibodies
 CC which are useful for detecting cancer. Zgpai polypeptides having
 CC anti-microbial activity are useful for treating dental caries,
 CC periodontal disease, thrush, gastrointestinal disease, urinary tract
 CC infections, respiratory infections, vaginal infections, acquired immune
 CC deficiency syndrome (AIDS) and lung infections associated with cystic
 CC fibrosis and prevention of infection in skin and other epithelial wounds.
 CC Zgpai-cytokine fusion proteins are useful for enhancing in vivo killing
 CC of target tissues (epithelial cancers, and more specifically lung,
 CC ovarian and rectal cancers). Zgpai polypeptides, fragments, fusion
 CC proteins or agonists are useful in in vitro studies of exogenous
 CC microorganism infections such as bacterial, viral or fungal infection and
 CC also to study epithelial cell defense in induction in cell culture. Zgpai
 CC antibodies, polynucleotides and polypeptides are useful for detection of
 CC zgpai polypeptide, mRNA or anti-zgpai antibodies, thus serving as markers
 CC for detecting genetic diseases or cancers. Zgpai sequences are useful as
 CC diagnostics in forensic DNA profiling and as probes or primers to clone
 CC 5' non-coding regions of a zgpai gene. Zgpai antibodies are useful for
 CC tagging cells that express zgpai, for screening expression libraries and
 CC as neutralizing antibodies or as antagonists to block zgpai activity
 CC in vitro and in vivo. Zgpai gene is also useful in gene therapy.
 XX
 XX SQ Sequence 1110 BP; 249 A; 319 C; 295 G; 247 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,286-33 Length: 1110
 Score: 413.00 Matches: 76
 Percent Similarity: 62.11% Conservative: 42
 Best Local Similarity: 40.00% Mismatches: 66
 Query Match: 39.30% Indels: 6
 DB: 22 Gaps: 3
 US-09-462-625-4 (1-191) x AAD04006 (1-1110)
 QY 4 AlATrPAlaLeuProSerLeuLeuArg-----LeuGlyAlaIleGlnGluThr 19
 DB 541 AGTTATGTCAGCCACTTCTGTGAAAGGAGAGACTGCTGCGCCCTCGCAGAGACA 600
 QY 20 GLUASPRoAlAcYsCySeSerProIleValProArgAsnGluTrpIlyAlaLeuAlaSer 39
 DB 601 AGCCTGAAGAAGGCTGCGCCGCGCTTCCACGCGTCTGTGTGGGAGCCAGGAGAC 660
 QY 40 GLUCysAlaGlnHisLeuSerLeuProLeuArgTyrValValSerHisThrAlaGly 59
 DB 661 CACTGTCCAGG---ATGACTCTCCAGGAGATGATGCGCATTAATCCACACTCCGGG 717
 QY 60 SerSerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisTyrHis 79
 DB 718 AGGACCTGCAACATTCTGTATGATGAGTCCGCTGTGTCGGGACATCCAGTCTTCTAC 777
 QY 80 MetLysThrLeuGlyTyrCysAspValGlyTyrAsnPhenLeuIleGlyGluAspGlyLeu 99
 DB 778 ATAGACAGGCTCAAGCTATGCGACATGTGTTATTAATCTCTGTGTGGGACAGATGGCGCC 837
 QY 100 ValTyrGlnGlyArgGlyTyrAsnPhenThrGlyAlaHisSerGlyHisLeuTyrAsnPro 119

Dd		838	ATTTATGAGGGGTGGCGTGATGTCACAGCCTCC---TCCACCCTGGCTFACGATGAC	894
Oy		120	MettSerIleGIYtLeserPheMetGIyaNtYrMetaspargValProthrProglNaIa	139
Dd		895	ATTGCCCTGGCATTTACTTWCATGAGGCACCTTCACAGATATACACCAATGCTGCAGCA	954
Oy		140	IleAArgAlaIalagInclYleuleuAlaIacysGIYalaIaInclYlaIaleuArSeraSn	159
Dd		955	CTTAGAGGCCGCCMAACACTGATCAGTGTGCCATGTGTCAAAGGTCCTGACTCCCAAC	1014
Oy		160	TyrValIleuLySGlyHisArGPaspValaGlnaRgfHriueSerProGIYaSnGlnLeuTyf	179
Dd		1015	TACCTGCTGTGGTGGCCACAGTGAATGTGGCCCCGAACTGTCTCTCTGGCAGGCTTTGAC	1074
Oy		180	HISleuIIeGLAsnTrpPROHIStyrArg	189
Dd		1075	AACATCATCAGCACCGCTGGCTCATTTCCAAM	1104
RESULT 15				
ABA90337				
ID		ABA90337	standard; CDNA; 1110 BP.	
XX		AC	ABA90337;	
XX		AB		
DT		12-FEB-2002	(first entry)	
XX		DE		
XX		DE	Human polynucleotide #12.	
KW		Human; nocotropic; neuroprotective; anticonvulsant; antidepressant;		
KW		neuroleptic; tranquiliser; antiarrhythmic; cardiant; antiasthmatic;		
KW		antiinflammatory; antiliphaemic; hepatocrotropic; virucide; antidiabetic;		
KW		neptrocrotic; anorectic; cytosaltic; vaccine; neurological disease;		
KW		cardiovascular disease; respiratory disease; liver disease;		
KW		renal disease; skeletal muscle disease; gastrointestinal disease;		
KW		placental disease; testicular cancer; male fertility; pancreatic disease;		
KV		ss.		
OS		Homo sapiens.		
PN		WO200181363-A1.		
XX		01-NOV-2001.		
XX		26-APR-2001;	2001WO-US13360.	
XX		27-APR-2000;	2000US-199963P.	
PR		11-MAY-2000;	2000US-203336P.	
PR		25-MAY-2000;	2000US-207087P.	
PR		26-MAY-2000;	2000US-207546P.	
XX				
PA		(SMIK) SMITHKLINE BEECHAM CORP.		
XX		(SMIK) SMITHKLINE BEECHAM PLC.		
PI		Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;		
PI		Lai Y, Xie Q;		
DR		MPJ; 2002-041392/05.		
DR		P-PSDB; ABB53272.		
XX				
PT		Novel polypeptides and polynucleotides useful as a vaccine for		
PT		preventing and treating diseases associated the polypeptide, e.g.		
PT		Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,		
PT		asthma, amnesia -		
XX				
PS		Claim 2; Page 51; 116pp; English.		
XX				
CC		The invention relates to an isolated polypeptide comprising a 277, 480,		
CC		583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,		
CC		844, 782, 262, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,		
CC		784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as		
CC		given in the specification. The polypeptides, modulators of the		
CC		polypeptides and antibodies against the polypeptides are useful for		

treating diseases such as neurological and psychiatric diseases
including Alzheimer's, paramyotonic nuclear palsy, Huntington's disease,
myotonic dystrophy, anorexia and depression; cardiovascular diseases
including congestive heart failure, Hodgkin's disease and myocardial
infarction; respiratory diseases including asthma, chronic obstructive
pulmonary disease, cystic fibrosis and adult respiratory distress
syndrome; liver diseases including hypercholesterolemia, cirrhosis,
viral and nonviral hepatitis, type II diabetes mellitus, and impaired
glucose tolerance; renal disease including renal failure, acute tubular
necrosis and glomerulonephritis; skeletal muscle diseases including
Eulenburg's disease, hypoglycemia and obesity; gastrointestinal
diseases including myotonic dyspraxia; diseases of placenta including
choriochorionoma; diseases of testes including testicular cancer,
male reproductive diseases including low testosterone and male
infertility, and disease of pancreas including diabetic ketoacidosis,
type I and 2 diabetes and obesity. The present sequence encodes a
polypeptide of the invention.

```

SQ Sequence 1110 bp; 249 A; 319 C; 295 G; 247 T; 0 other
Alignment Scores:
Pred. No.: 2,28e-33 Length: 1110
Score: 413.00 Matches: 76
Percent Similarity: 62.11% Conservative: 42
Best Local Similarity: 40.00% Mismatches: 66
Query Match: 39.30% Indels: 6
DB: 24 Gaps: 3
US-09-462-625-4 (1-191) x ABR90337 (1-1110)

```

US-09-462-625-4 (1-191) X ABA90337 (1-1110)

```

Oy      4  AlaThrPalaIleuProSerIleuLeuArg-----LeuGlyValAlaGlnIleuThr 19
Db      541 AGTATTGTTCAAGCACTTCTTGTAAGAGCGAGAACTGCTGCTGGCCCTGGGAGAAACA 600
Oy      20  GluAspProAlaCysCysSerProIleValProArgAsnGluTyrLysAlaLeuAlaSer 39
Db      601 AGCCTGAAGAAGGCTTGCCCCGCTTGTCGCCACGCTGTGTGGGGAGGCCAGGGAGACC 660
Oy      40  GluCysAlaGlnIleuSerLeuProLeuAlaGlyTyrValAlaValSerIleThrAlaGly 59
Db      661 CACGTGCTCCACGG--ATGACTCTCCACGAGAAAGTAAGCATATTATCCACACTGCCGG 717
Oy      60  SerSerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValAlcIleIleThrHis 79
Db      718 AGGACCTCGCAACATTTCTGATGATGCTGCCCTGCTGCTGGGACATCCAGCTTTCTTAC 777
Oy      80  MetLysThrLeuGlyTyrCysAspValAlaGlyTyrAsnPheIleGlyValAspLysLeu 99
Db      778 ATATACACGGCTCAAGTCATCATCGCATGGTGTATTAATCTCTGTGGCCAGGATGGCCCC 837
Oy      100  ValTyrGluGlyArgGlyTyrAsnPheThrGlnAlaHisSerGlyHisLeuThrAsnPro 119
Db      838 ATTATATAGAGGGGTGGGCTGGATATGTCCAAAGCTCC--TCCACCCCTGGCTACATGAC 894
Oy      120  MetSerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAla 139
Db      895 ATATCCCTGGGCAATTACCTTCATGGGCACTTCACAGGTATCCACCACCAATCTGCACACA 954
Oy      140  IleArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsn 159
Db      955 CTAAAGGCAACCCCAAGACTCGATTCACAGTGTGCCATGTGTAAAGGTACTGACTGCCAAC 1014
Oy      160  TyrValLeuLysGlyHisArgAspValAlaGlnArgThrLeuSerProGlyHisAsnGlnLeuTyr 179
Db      1015 TACCTGCTGGTGGGCGCACAGTGAATGTGGGCCCGCAACCTTGTCTCTGGGACAGGCTTTGTAC 1074
Oy      180  HisLeuIleGlnAsnTyrProHisTyrArg 189
Db      1075 AACATCATCAGCACACTGGCTCATTTTCAAA 1104

```

Search completed: November 12, 2002, 04:15:43

Wed Nov 13 09:37:10 2002

us-09-462-625-4.rng

Page 14

Job time : 317 secs
